Scuola Internazionale Superiore di Studi Avanzati

PhD course in Theory and Numerical Simulation of Condensed Matter

---

COLOR OPTICAL PROPERTIES OF ANTHOCYANINS IN SOLUTION

A multi-scale computational study

A thesis submitted for the degree of *Doctor Philosophiae*

Supervisors:
Prof. Stefano Baroni
Dr. Alessandra Magistrato

Candidate
Mariami Rusishvili

The research presented in this thesis has been part of a five-years collaboration between SISSA and *Mars Wrigley Confectionery*. I am particularly grateful to Drs. Rebecca Robbins and Tom Collins of the Mars R&D team for frequent, regular, and enlightening discussions on the planning and results of my research.

Contents

Chapter 1

Introduction

Colors are appreciated both for their intrinsic esthetic value and as a basis for identification and quality judgment. Thus, they play a very important role in our lives, whether we realize it or not. They have the ability to affect our emotions and moods in a way that few other things can. Being a complex subject with many strands, it has the power to subliminally convey values and stories. In the food industry visuals are growing in importance today, and the future looks rosy for brighter, simple colors that are both functional and attention-getting.

Historians of consumer culture have shown how an increasing number of firms began in the early twentieth century to capitalize on color for various commodities, including clothes and automobiles. For these products, color variety became a crucial element of brand identity and consumer choices. The color of foods cannot, however, be understood solely as an indicator of abundant varieties. The calibration of color was essential for food manufacturers to designate flavor, prevent the discoloration of foods, and make foods look natural and fresh. Food coloring has been a common practice across cultures for millennia, at least since ancient Egyptians used saffron to color various foods. Before the introduction of synthetic dyes in the late nineteenth century, so-called natural dyes derived from plants and organic minerals, including saffron, indigo, and cochineal, had been the major source of coloring foods in many parts of the world. The use of these dyes was limited as they were expensive. Due to the high prices on them industries started to exploit the artificial colorants starting from the end of 19th century [1, 2]. However, in the past couple of decades, concerns about the safeness and healthiness of those colorants started to arise. The apprehension and distrust of the public is understandable in view of the

recent delistings of several food dyes previously enjoying official approval as 'safe'. Some of the colorants that are synthetic compounds not naturally present in foods are even suspected to be allergenic, carcinogenic, or to have potentially adverse behavioral effects on children predisposed to hyperactivity and attention-deficit disorders [3]. Because of the legal prohibitions on some colorants the interest in natural dyes spurred. The most common natural pigments responsible for some of the most spectacular natural colors are anthocyanins [4, 5]. The purpose of this thesis is to unveil the molecular and geometrical mechanisms responsible for the optical properties of anthocyanins.

Figure 1.1: Some examples of fruits, vegetables and flowers that are rich in anthocyanins

1.1 Anthocyanins

The word anthocyanin coined in the beginning of the 19th century. It is derived from the Greek words *anthos* meaning *"flower"* and *kyaneos/kyanous* meaning *"dark blue"* to designate the blue pigments of flowers. Later it was understood that not only blue but all the colors appearing on flowers, fruits, leaves, stems, and roots (except tomatoes and red beets) in a red-purple-blue gamut are attribute to pigments chemically similar to the one originally called anthocyanin. Anthocyanins impact esthetic appeal and serve as quality indicators in fruits and vegetables. As an example when sufficient anthocyanin has appeared in the skin of red grapes they are considered to be ripe.

Interest in anthocyanins has been increased as a result of the continuing official delisting of artificial food dyes [6–10]. Anthocyanins are considered to be the best replacements for the banned dyes because of their abundance in the nature and their wide range of colors. Apart from that, they are known to be safe since people have been consuming them for countless generations without having any side effects. One more advantage of anthocyanins is that they are water-soluble phytochemicals which makes it easier

to incorporate them into aqueous food systems. On the other hand, they are characterized as pigments with very poor stability compared to artificial colorants. During the past century anthocyanins have been extensively investigated from chemical, biochemical and food technological points of view. Despite the interesting developments in this field, a lot of additional information is needed about their chemical constitution and stability. A deep understanding of chemical and geometrical mechanisms responsible for optical properties and stability of these dyes would therefore be crucial to allow industry the successful usage of anthocyanins as colorants in food.

There is an enormous variety of anthocyanins spread in nature (names and abbreviations of the most common are shown on 1.3). The abundance and variety makes them an interesting group to study. Anthocyanins have the same bio-synthetic origin as other natural flavonoid compounds and, like other flavonoids, are structurally characterized by the presence of a carbon skeleton made up of a $C_6C_3C_6$ skeleton unit [2]. However, unlike other flavonoids, anthocyanins strongly absorb visible light and thereby generally confer an infinite variety of colors [1, 4, 11]. The main part of anthocyanins is its aglycone, the flavylium cation (see figure 1.2) (backbone). The backbone is made of conjugated A and C double and B single rings, that are responsible for the absorption of light. The aglycones with their substituents at positions (3,5,7,3′,4′ penta) or (3,5,7,3′,4′,5′ hexa) are called anthocyanidins. Up to now there are reports of more than 500 different anthocyanins which are derived from 23 anthocyanidins. Among those 23 anthocyanidins only 6 of them (Pelargonidin, Cyanidin, Delphinidin, Peonidin, Petunidin, Malvidin see figure 1.4) are the most common from the food point of view. The above mentioned 6 species differ in the number of hydroxyl and methoxyl groups attached in the B ring of the flavylium cation at positions 3′ and 5′. Anthocyanidins are quite unstable in solution by themselves, that is why they mainly occur in glycosylated forms which makes them more friendly to the water solution. Anthocyanidins with one or more glycosyl units are called anthocyanins. Glycosylation is possible at positions 3,5,7,3′,4′ and 5′ and anthocyanidins can be mono/di/tri glycosylated (anthocyanidins with one, two or three glycosyl units) [1,2]. The most common glycosyl units of anthocyanins found in various fruits, vegetables and flowers are shown in figure 1.5.

Anthocyanins can be acylated as well. Usually in case of acylated anthocyanins, organic acids are attached to anthocyanin glycosyl units through

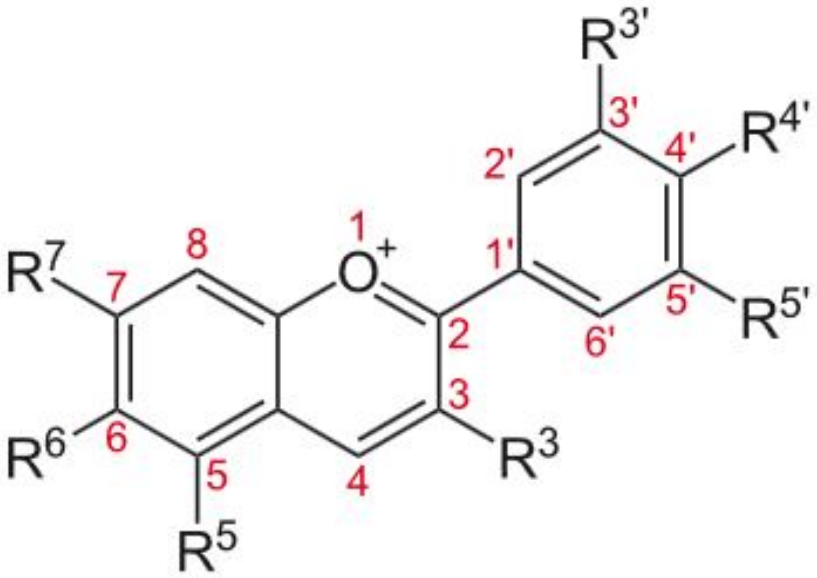


Figure 1.2: The flavilium ion, the basic structure (backbone) of anthocyanins

	Substitution pattern							
Anthocyanidin	3	5	6	7	3´	4´	5´	color
Carajurin	H	H	OH	OH	H	OCH_3	OCH_3	-
Arrabidin	H	H	OH	OH	H	OH	OCH_3	-
3´-Hydroxyarrabidin	H	H	OH	OH	OH	OH	OCH_3	-
Apigenin	H	OH	H	OH	H	OH	H	Orange
Luteolin	H	OH	H	OH	OH	OH	H	Orange
Tricetinidin	H	OH	H	OH	OH	OH	OH	Red
Pelargonidin	**OH**	**OH**	**H**	**OH**	**H**	**OH**	**H**	Orange
Aurantinidin	OH	OH	OH	OH	H	OH	H	Orange
Cyanidin	**OH**	**OH**	**H**	**OH**	**OH**	**OH**	**H**	Orange red
5-Methylcyanidin	OH	OCH_3	H	OH	OH	OH	H	Orange red
Peonidin	**OH**	**OH**	**H**	**OH**	**OCH_3**	**OH**	**H**	Red
Rosinidin	OH	OH	H	OCH_3	OCH_3	OH	H	Red
6-Hydroxycyanidin	OH	OH	OH	OH	OH	OH	H	Red
6-Hydroxydelphinidin	OH	OH	OH	OH	OH	OH	OH	Bluish red
Delphinidin	**OH**	**OH**	**H**	**OH**	**OH**	**OH**	**OH**	Bluish red
Petunidin	**OH**	**OH**	**H**	**OH**	**OCH_3**	**OH**	**OH**	Bluish red
Malvidin	**OH**	**OH**	**H**	**OH**	**OCH_3**	**OH**	**OCH_3**	Bluish red
Pulchellidin	OH	OCH_3	H	OH	OH	OH	OH	Bluish red
Eupinidin	OH	OCH_3	H	OH	OCH_3	OH	OH	Bluish red
Capensinidin	OH	OCH_3	H	OH	OCH_3	OH	OCH_3	Bluish red
Hirsutidin	OH	OH	H	OCH_3	OCH_3	OH	OCH_3	Bluish red

Figure 1.3: Main groups of anthocyanidins. R1-R7 are the side groups shown in upper figure

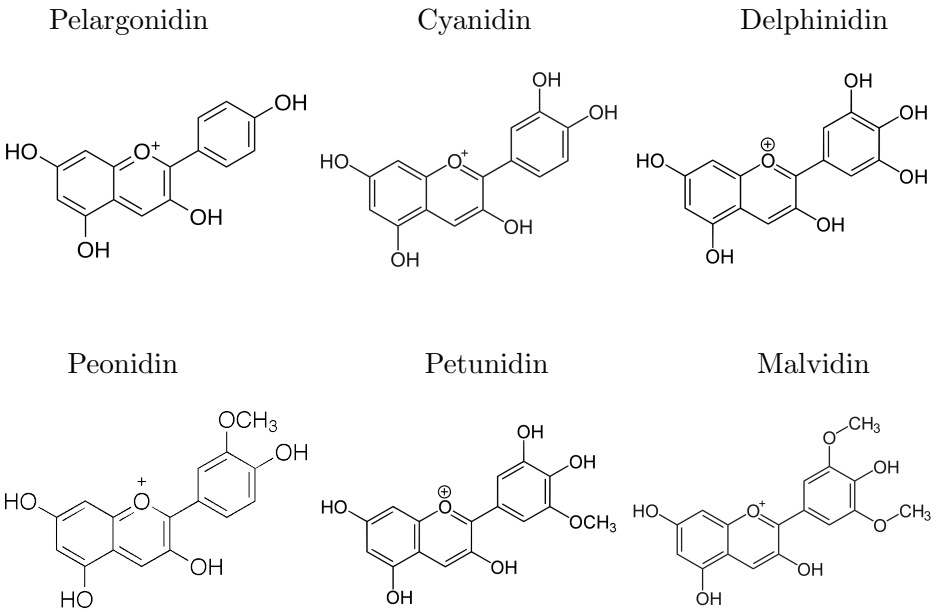


Figure 1.4: The most common anthocyanidins from which more than 90% of anthocyanins found in nature are derived.

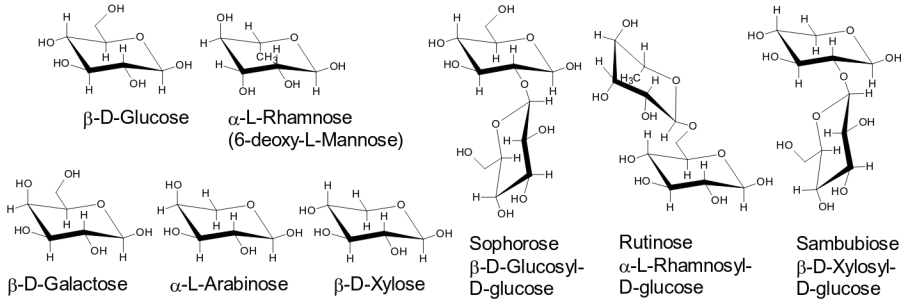


Figure 1.5: The most common glycosyl units attached to the anthocyanidins that are found in nature.

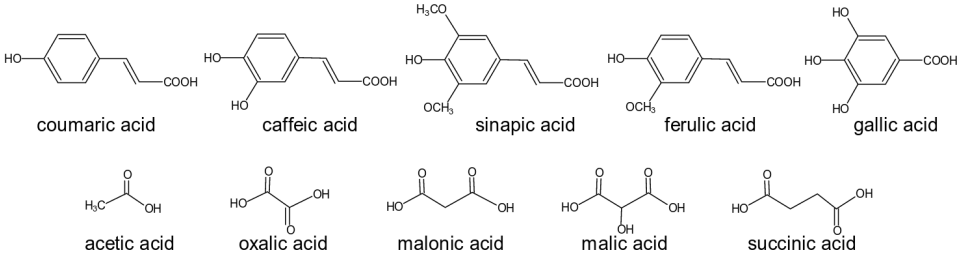


Figure 1.6: The most common acyl units attached to the anthocyanins that are found in nature.

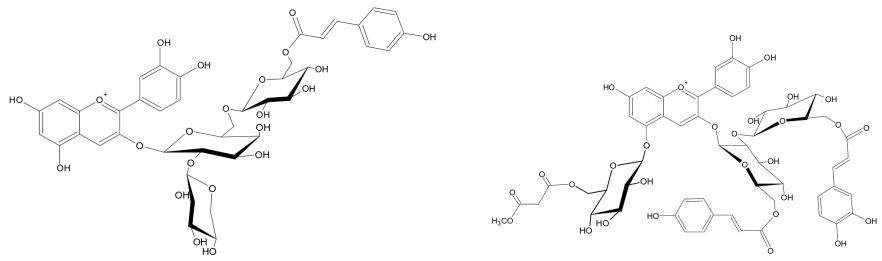


Figure 1.7: Some examples of anthocyanins with complex substituents. Left: Cyanidin 3-(2"-xylosyl-6"-(6-coumaroyl)-glucosyl)-galactoside; Right: Cyanidin 3-((2"-(6-caffeoyl)-glucosyl)-6"-(coumaroyl)-glucoside)-5-(6"-(metylmalonyl)-glucoside.

an ester bond. Commonly, anthocyanins can be attached by aromatic phenolic acids, aliphatic dicarboxyl acids or a combination of both of them.

The optical properties of anthocyanins are determined by all which is mentioned above: different patterns of hydroxyl and methoxyl groups attached at the B ring, sugars, and acylated sugars change the absorption spectrum of the molecule. Apart from these structural effects the optical properties of anthocyanins and consequently the color expressed by them is strongly influenced by the physicochemical milieu in which they are viewed. The same anthocyanin may express different colors, depending on the pH and concentration of the solution, the presence of copigments, metal complexations and other environmental conditions (temperature, light...).

As mentioned above the color of anthocyanins is very sensitive to the acidity of the solvent [1, 2]. The structure of anthocyanins changes significantly with changing pH [2]. At acidic conditions pH < 2 the only stable form present in solution is the flavylium cation (AH^+). While increasing the pH value, deprotonation of the hydroxylated flavylium cation occurs. Depending on how many hydroxyl groups are presents (at positions 4′,5 and 7) where proton loss can occur we may end up with two or three tautomeric species that are probable to coexist at slightly acidic conditions (pH 4-6). The six most common anthocyanins mentioned above (see figure 1.4) have all three ionizable hydroxyls, so three neutral prototropic tautomers (quinonoidal bases) $A^{\circ}_{4'}$, A°_{7}, A°_{5} can be formed after the proton loss (see figure 1.8). The concentration of each tautomeric species strongly depends on the type of anthocyanin and can be evaluated by studying the free energy surface of the proton transfer reaction between tautomeric species (refer-

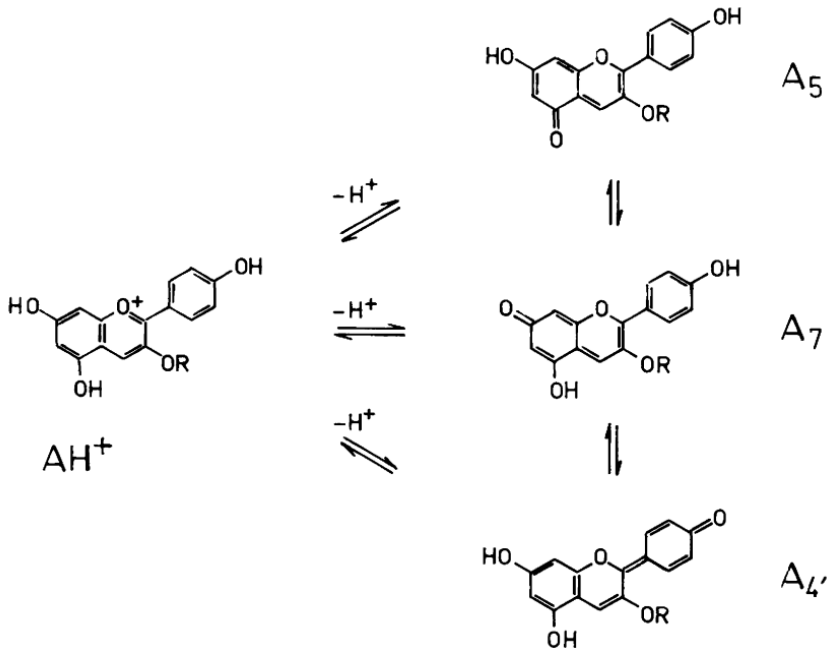


Figure 1.8: At slightly acidic conditions the flavylium cation to loses a proton from a free hydroxyl group. This figure shows a case in which three free hydroxyl group are available and deprotonation at each position is possible. In this case we may end up with three tautomeric states at the same pH solution. Of course each of those tautomeric states will have its probability of existence depending on the anthocyanin.

ence sara). This is possible since it is known that inter-conversion from one tautomer to another does not necessarily occur via the flavylium structure. Proton transfer can happen from one protonated hydroxyl group to a deprotonated one in water solution [2]. When increasing the pH value further deprotonation is possible if more than one hydroxyl group was available in the initial flavylium case. After a second deprotonation negatively charged species are formed. When two hydroxyl groups are still present in the neutral tautomers the situation is highly complex and is fully represented on the figure 1.10. Each neutral tautomeric species is able to deprotonate in two different positions. $A^{\circ}_{4'}$ can deprotonate at positions 5 or 7 and can form $A^{-}_{4'5}$ or $A^{-}_{4'7}$, A°_{5} can deprotonate at positions $4'$ or 7 and can form $A^{-}_{4'5}$ or A^{-}_{57}, A°_{7} can deprotonate at positions $4'$ or 5 and can form $A^{-}_{4'7}$ or A^{-}_{57}. This results in three tautomeric states for negative anthocyanins. The same analysis as discussed above for understanding the relative populations of each species of neutral states applies to negative states as well. When

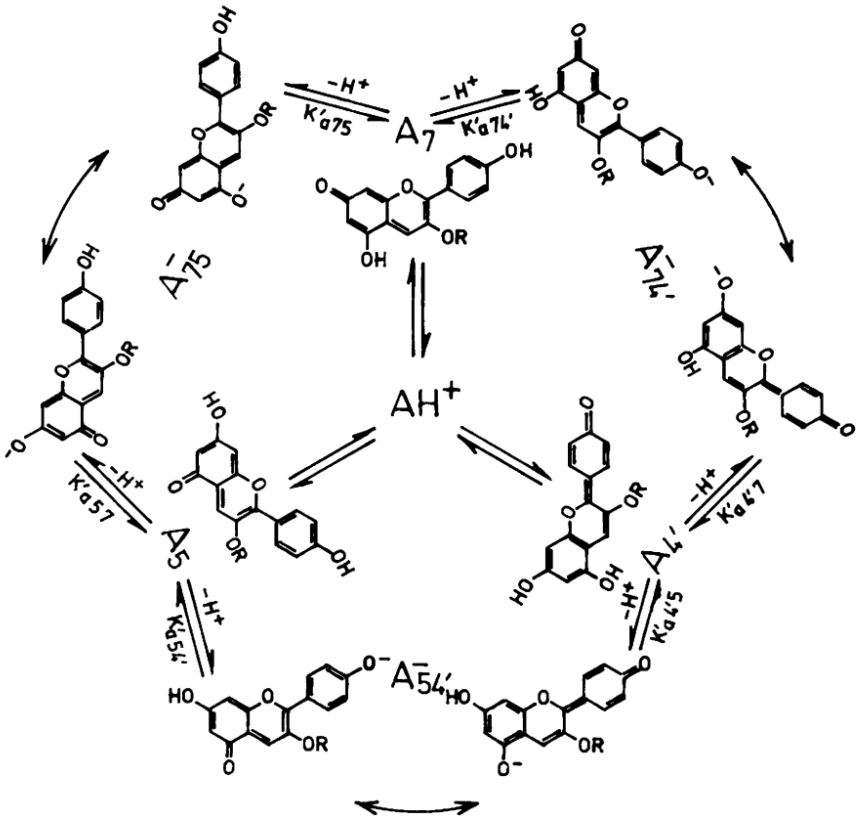


Figure 1.9: This is the continuation of the deprotonation pattern from neutral species of anthocyanins to the negative ones, when in the initial flavylium case three free hydroxyl groups where present. From $A_{4'}$ after deprotonating at positions 5 or 7 $A^-_{4'5}$ or $\mathsf{A}^-_{4'7}$ negative tautomeric states are formed, from A°_5 deprotonating at positions 4′ or 7 $A^-_{54'}$ or A^-_{57} are formed, and from A°_7 deprotonating at positions 4' or 5 $A^-_{74'}$ or A^-_{75} are formed. Stating that $A^-_{4'5}$ and $A^-_{54'}$ are the same species as $A^-_{4'7}$ and $A^-_{74'}$; and as A^-_{57} and A^-_{75} we can state that we are ending up with three different tautomeric states of negative anthocyanins from now on referred to as $A^-_{4'7}$, $A^-_{4'5}$ and A^-_{57} [2].

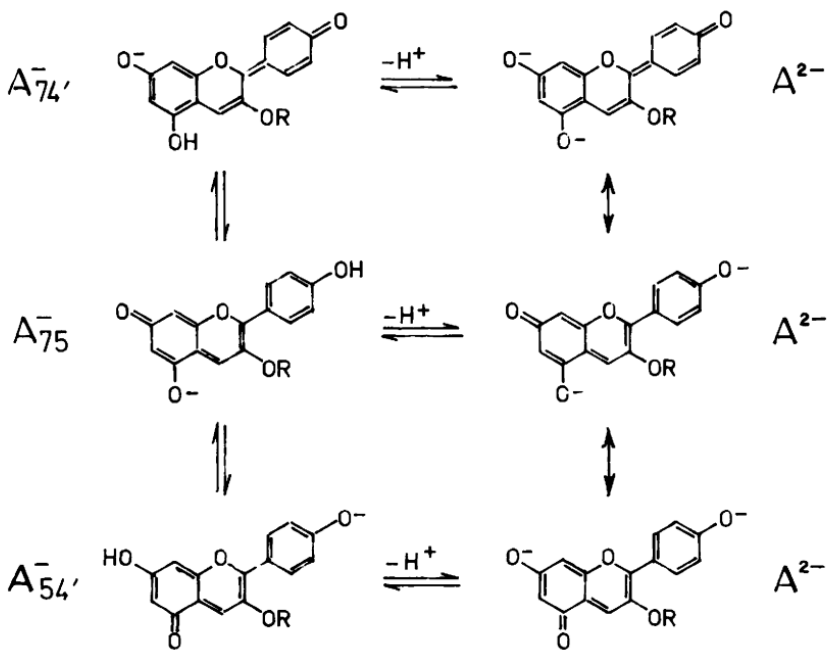


Figure 1.10: When the medium is sufficiently alkine all the negative tautomeric states lose one last proton. $A^{-}_{4'7}$ loses proton from the position 5 and forms $A^{2-}_{4'75}$, $A^{-}_{4'5}$ loses proton from the position 7 and forms $A^{2-}_{4'57}$, and A^{-}_{57} loses proton from the position 4′ and forms $A^{2-}_{574'}$. We can see that all three $A^{2-}_{4'75}$, $A^{2-}_{4'57}$, $A^{2-}_{574'}$ states are equal and correspond to the state when initial flavylium cation is deprotonated at positions 4′,5 and 7. So at alkaline pH we end up with a single dianion state from now on referred as A^{2-}. [2]

initially three hydroxyl groups were present at positions 4′, 5 and 7 in the flavylium form, and moving to sufficiently alkaline conditions, one more deprotonation can occur and from any of the negative tautomeric species $A^{-}_{4'7}$, $A^{-}_{4'5}$, A^{-}_{57} a single dianion A^{2-} state is formed by losing a third proton. This deprotonation scheme is represented in figure 1.10.

To sum up, the structure of anthocyanins undergoes significant changes by varying the acidity. At strongly acidic conditions the only stable form is the flavylium cation (AH^{+} on the figure 1.11), at slightly acidic or neutral conditions quinonoidal bases (A on the figure 1.11) exists in different tautomeric forms, the number of tautomers depending on the hydroxide group pattern on the B ring of the chromophore. At pH values from 4 to 6 exist another more stable form called carbinol (B on the figure 1.11), which in practice accurs when position 2 of the chromophore is fully hydrated. In this form the anthoyanin does not have its conjugated double

bond between the double AC and single B rings, and consequently it does not absorb in the visible range, making this species colorless. The transition from carbinol to flavylium state is reversible, but carbinol can also transform to a chalcone pseudobase (C on the figure 1.11), where the C ring of the chromophore is open. This last transition is irreversible and the chalcone forms are colorless as well. On figure 1.11 the ratio of AH^+, A, B and C forms is expressed as a function of the pH for the malvidin-3-glucoside. The degree of coloring of a solution at a given pH is determined by the concentrations of the coexisting stable forms at that pH. Experimentally it is known that malvidin-3-glucoside does not confer much color to a solution in which pH ranges from 4 to 6. It is also confirmed from the equilibrium distribution showed on figure 1.11, where at those pH values, the proportion of colorless carbinol species dominate in solution. For pH values ranging from 6 to 9 intensively colored solutions are produced. The species responsible for color at those acidic conditions are mainly the negatively charged A^-.

In the late 50s [12, 13] anthocyanins extracted from a blue flower were found to contain Mg. Moreover, it was found that the complex composed of metal ions, anthocyanins and flavones(backbone of 2-phenylchromen-4-one (2-phenyl-1-benzopyran-4-one)), had stoichiometric amounts of each of them and this complex was called metalloanthocyanin. The composition is fixed at 6:6:2 ratio of anthocyanins, flavones and metal ions respectively [4]. Lately it was discovered that Mg is not the only metal that can form complexes with anthocyanins and that 1:1 metal ANT complexes are also very probable in solution. The other metals are Ca, Al, K, Fe and the position of binding is strongly dependent on the hydroxyl pattern on the B ring. The structures containing metal ions appear to be very stable in a wide range of pH and their color is shifted toward the blue end of the spectrum. One of the issues with the usage of anthocyanins as colorants is to achieve standardized color and increased stability. It seems that the complexation of anthocyanins with metal ions could be a solution to this problem. The possibility of getting bright and stable colors is the motivation for studying the impact of metal complexation on the behavior of anthocyanins in term of electronic, dynamical and optical properties.

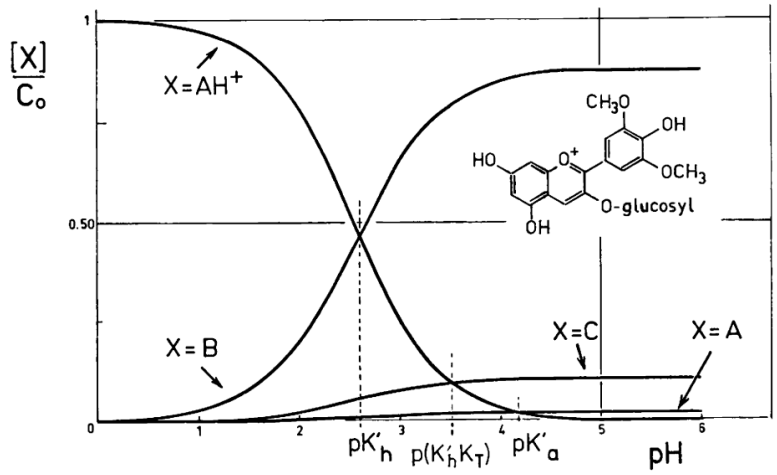


Figure 1.11: The distribution of the different Malvidin - 3 O glucoside equilibrium forms according to pH [2].

1.2 Plan of the thesis

In this work we aim to elucidate the main features affecting the optical properties of anthocyanins, their complexes with metals and some aspects of copigmentation, by computing the absorption spectra in various conditions. These spectra and the resulting perceived color of a pigment in solution depend on the different out-of-equilibrium molecular geometries that are reached through thermal fluctuations, as well as on the interaction with the solvent. For this reason we are introducing a newly developed multiscale protocol that accounts for each relevant effect at the appropriate time/length scale, through a combination of different computer simulations, resulting in computed absorption spectra which are comparable with experimental spectra.

This thesis is organized in the following way: In chapter 2 the theoretical background of this work is introduced. Firstly, the problem and the computational challenges are described, followed by the separate ingredients needed for our multiscale computational protocol. In chapter 3, this protocol is used to study the optical properties of different charge states of cyanin-3-glucoside. In chapter 4 we study the role of metal ions on the absorption spectrum when they form complexes with anthocyanins. Finally, in chapter 5 intramolecular copigmentation and its effect on optical properties are studied.

Chapter 2

Theoretical Background

In this chapter we illustrate the theoretical background for the ingredients used in our multiscale computational study of the absorption spectrum of complex molecules. We have combined different classical and quantum mechanical methods to accurately dissect the structural, dynamical, electronic, and optical properties of large biological molecules. We account for the conformational equilibrium of anthocyanins in explicit water by identifing the free energy basins using microsecond-long enhanced sampling molecular dynamics (MD) simulations [14] (sections 2.1 and 2.2). We use a clustering algorithm [15] (appendix 2.3) to investigate the conformational landscape of the selected systems. To consider the effect of thermal fluctuations and the intermolecular interaction with explicit water molecules, we employ ab initio MD simulations [16] (section 2.6) in explicit solvent on the selected conformers (from the clustering algorithm), as extracted from classical MD. Finally, time dependent density functional theory (TDDFT, section 2.5) is used on selected independent frames extracted from the ab initio trajectory considering a droplet of strongly hydrogen (H)-bonding water molecules and a dielectric continuum model [17] to compute the absorption spectrum (see figure 2.1).

Optical absorption occurs over a much faster time scale than thermal fluctuations, whereas the latter are much faster than human perception. Therefore, the perceived color of a solution results from the time average of the molecular spectra corresponding to different time-dependent molecular geometries, $\{\mathbf{R}(t)\}$, as:

$$\kappa(\lambda) = \frac{1}{T}\int_0^T \kappa\big(\lambda, \mathbf{R}(t)\big)\, dt, \tag{2.1}$$

where $\kappa(\lambda, \mathbf{R})$ is the absorption coefficient of the solution as a function of the wavelength of the light and $\mathbf{R}$ indicates the set of all the molecular coordinates. In Ref. 17 is presented a multi-model approach (developed in our group) to evaluating the average spectra in Eq. (2.1). This approach is based on a combination of explicit-solvent AIMD to sample the molecular configurations, where the solvent is treated with atomic resolution, with an implicit-solvent approach to TDDFT to compute the molecular spectra along the MD trajectory, where the solvent is mimicked by a polarisable continuous medium. Even the simplest anthocyanin in solution is characterized by a complicated configurational space. Its different conformational isomers are separated by free-energy barriers that are too high to be overcome on the time scale accessible to AIMD (a few dozen ps). A proper evaluation of the time averages in Eq. (2.1) would thus require such long MD runs that can only be afforded using classical force fields. Moreover, enhanced MD is required to access all free energy minima in a reasonable amount of simulation time. As molecular spectra are rather sensitive to the details of the molecular geometry, it is essential that the latter is sampled with AIMD accuracy. In order to reconcile the conflicting requirements of an accurate sampling within each conformational free-energy basin and of a statistically meaningful sampling of different conformers, we break the integral in Eq. (2.1) into trajectory segments spanning a same conformer and rewrite it as:

$$\kappa(\lambda) = \sum_i p_i \left\langle \kappa\big(\lambda, \mathbf{R}\big)\right\rangle_i, \tag{2.2}$$

where the index i labels different conformers, p_i is the relative population of each of them, and the symbol $\langle\cdot\rangle_i$ indicates a time average restrained to the i-th conformer. In our multi-scale protocol the p_i's are estimated from the relative residence times resulting from a long classical MD simulation, whereas the intra-conformer averages, $\langle\cdot\rangle_i$, are performed by evaluating the optical spectrum on the fly over shorter AIMD trajectories generated independently for each conformer. In order to identify the relevant conformers,

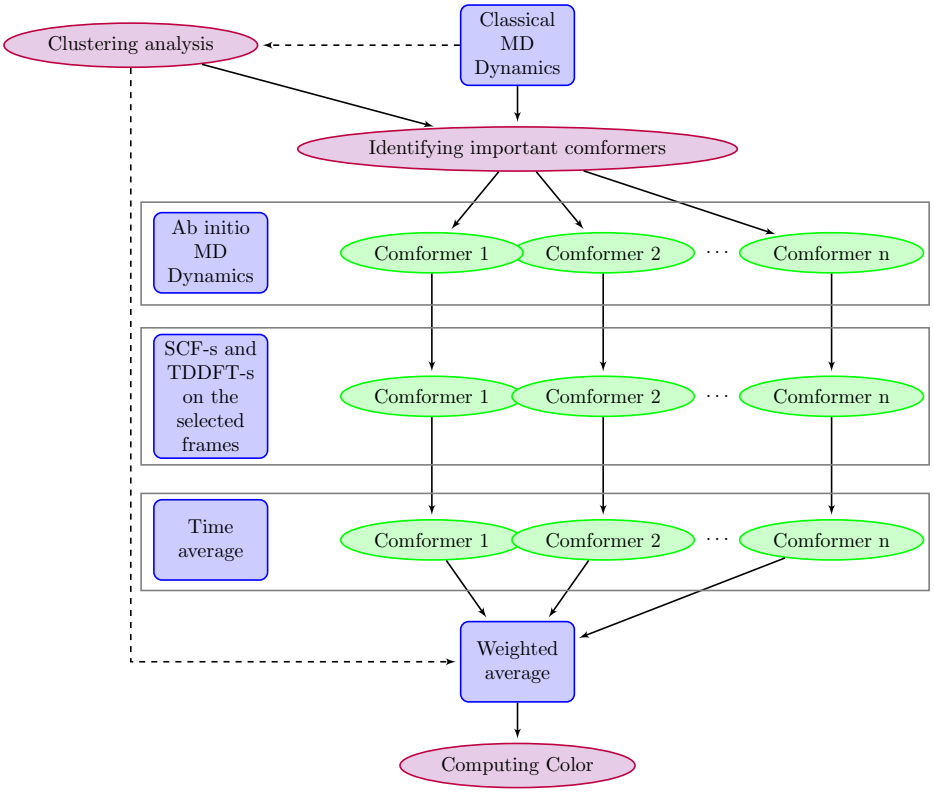


Figure 2.1: The scheme representing the multiscale computational protocol.

we make use of a clustering technique, based on the analysis of the statistical properties of a number of properly chosen slow variables – dihedral angles.

2.1 Molecular Dynamics

Computer simulations of biomolecular systems have gained a lot of popularity over the past few decades, passing from simulating very small systems in vacuo to large realistic solvated molecules, using molecular dynamics simulations [18–20]. Molecular Dynamics (MD) is a technique where the Newton's equations of motion are solved for the trajectory of N (classical) particles, that interact via a prescribed classical potential. The equations of motion for N particles has the following form:

$$m_i \ddot{\mathbf{r}}_i(t) = \mathbf{F}_i(t), \qquad i = 1...N, \tag{2.3}$$

with $\mathbf{r}_i$ being the atomic positions, m_i being atom masses and $\mathbf{F}_i(t)$ being the force on an atom i at the time t. The forces depend on the positions of all of the other $N-1$ particles at time t in the following way:

$$\mathbf{F}_i(t) = -\nabla_i v(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N, t) \tag{2.4}$$

$v(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N, t)$ being a potential.

$$\mathbf{F}_i(t+\Delta t) = -\nabla_i v(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N, (t+\Delta t)) \tag{2.5}$$

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t + \frac{1}{2m_i}\mathbf{F}_i(t)\Delta t^2 \tag{2.6}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{1}{2m_i}(\mathbf{F}_i(t) + \mathbf{F}_i(t+\Delta t))\Delta t \tag{2.7}$$

where $\mathbf{v}_i(t)$ it the velocity of particle i at time t.

At each step of dynamics the force on each particle is calculated according to the equation 2.5 and the positions and velocities of each particles are updated according to the equations 2.6 and 2.7 respectively.

For the first step of the MD simulations positions and velocities of each atom are required. The potential is also needed for computing the forces between each atom. In practice this potential has to be able to describe the Lennard-Jones interaction, it should also account for intra-molecular degrees of freedom like stretching or twisting, as well as describing many-body interactions, that can not be described with only a Lennard-Jones potential. This is why getting an accurate enough potential is a very challenging task. Most reliable potentials are derived by fitting some parameters to experimental data or to data coming from first principles calculations. However, there is also the problem of transferability, since the parameters are fitted to experimental or first principles data for a certain conditions. These set of parameters used to compute the potential energy are called force field (FF) parameters. It is not easy to guess how well the potential will perform while applied to other conditions. In order to get a good enough fit to ab initio results usually many expensive calculations are in general needed. So the choose of correct FF is crucially important for getting reasonable geometries during the MD simulations. Basically, a FF that can accurately describe big biological molecules does not exist, we are using FF generated by General AMBER Force Field (GAFF) [21] that seem to be accurate enough for anthocyanins.

2.2 Enhanced Comformational Sampling

Typical trajectories simulated with classical MD are nowadays on the order of 1μs, while conformational changes in molecules of the order of few tens of atoms need much longer time scales. Sometimes energy minima can trap the molecule for a long time and in this case the sampling process is very much slowed down, which leads to poor characterization of dynamical behavior. This issue has pushed the development of techniques that allow to effectively accelerate MD so that relatively low computational effort would allow us to explore all configurational space of the desired systems. Recently, a lot of advance sampling techniques were developed [22,23]. One class of methods (e.g. umbrela sampling [24–26] or metadynamics [27–30]) are based on biasing a small set of *a priori* chosen collective variables during the simulation. Metadynamics inserts memory in the sampling. By actually discouraging previously visited states be resampled, metadynamics allows one to direct computational resources to a broader exploration of the free energy landscape. Metadynamics can search through the entire free energy landscape, which makes it easier to study such slow phenomenons like protein folding, phase transitions and conformational changes. Another advantage of metadynamics is that it does not depend on a very accurate description of the potential energy surface being explored. Due to the fact that we discourage resampling of previous states, misevaluated conformations can be re-calculated and errors will tend to be "evened out". The hard task is to smartly choose the important degrees of freedom (collective variables) which sometimes can be a big problem, especially in case of large systems. Metadynamics does depend on a low dimensionality of the free energy landscape in order to produce an accurate description of the free energy surface, therefore using a small set of collective coordinates is essential. For huge biological molecules it would not be easy to describe the system with a small set of collective variables.

The other set of methods accelerates the configurational sampling by actually running separate replicas of the same system with different temperatures and exchanging the atomic positions every fixed time period with a Monte Carlo procedure. In a regular canonical molecular dynamics simulation, a configuration of one copy of the molecule under consideration is updated by molecular dynamics. On the other hand, in the class of methods like parallel tempering [31, 32] one considers an artificial system built up of N non-interacting copies of the molecule, each at a different temperature. The root of all these techniques lies in the simulated anneal-

ing procedure [32] a well-known method used in many optimization problems.Examples of parallel tempering methods are simulated tempering [33], or as above mentioned parallel tempering [31,34]. The issue of those methods id the choice of the number of replicas needed, which is correlated with the size of the system. When the free energy landscape is very rough, the preassigned temperature range grows and with it also the number of replicas needed. These limitations to handling "hardly-relaxing" systems in MD has led to the implementation of a new generalized-ensemble algorithm, the hamiltonian replica exchange method (HREM) [35–38]. In this method different replicas evolve not only at different temperatures but according to different Hamiltonians. One of the examples of HREM is Replica exchange with Solute Tempering (REST1) [39] which is higly applicable to large systems since the number of the replicas required is greatly reduced by heating up just the solute molecules while the solvent remains cold in higher temperature replicas. In REST1, the total interaction energy of the system is decomposed into three components: the solute intramolecular energy, E_{solute}; the interaction energy between solute and solvent, E_{ss}; and the self-interaction energy between solvent molecules, $E_{solvent}$. Replicas running at different temperatures evolve through different Hamiltonians involving relative scalings of the three components of interaction energy. The potential energy of the replica running at the temperature T_i has the following form:

$$E_i(X) = E_{solute}(X) + \frac{\beta_0 + \beta_i}{2\beta_i} E_{ss}(X) + \frac{\beta_0}{\beta_i} E_{solvent}(X)$$

where X represents the configuration of the whole system, $\beta_k = 1/k_B T_k$ and T_0 is the temperature of our interest. (while i=0 the potential reduces to the normal potential with all the scaling factors being 1).

However, it can happen that lower temperature replicas stay in one configuration and the high energy ones in another, and the exchange between those conformations is very low. Due to the fact that the thermodynamic properties of a system with potential energy E_k at temperature T_k, are the same as those for a system with potential energy $(T_0/T_k)E_k$ at temperature T_0, instead of using both different potential energies and different temperatures for different replicas, one can just run all the replicas at the same temperature albeit on different potential energy surfaces using Hamiltonian Replica Exchange Method called Replica Exchange with Solute Scaling [40]. While in REST1 enhanced sampling of the molecule conformations is achieved by increasing the temperature of the molecule

while exchanges between replicas were less probable due to high energy barriers, in REST2, enhanced sampling is achieved through scaling of the intra-molecular potential energy of the molecule by the factor of $\lambda = \beta_k/\beta_0$, and since this factor is always less than one, the barriers separating different conformations are lowered. Hence, each replica moves on a modified potential surface where the barriers between replicas in the intra molecular force field are reduced by the scaling. So in REST2, the differences between different replicas are not the temperatures of the replicas but the effective temperatures which in practice are accounted for by the different scaling factors used in describing potential energy. To be more specific, in REST2, all of the replicas are run at the same temperature T_0, and the potential energy for replica i can be written in the following way:

$$E_i(X) = \lambda E_{solute}(X) + \sqrt{\lambda} E_{ss}(X) + E_{solvent}(X) \tag{2.8}$$

The advantage of scaling the potential energy instead of the temperature is related to the fact that the energy is an extensive property, whereas the temperature is an intensive one. Hence, one can selectively choose a portion of the system and specific parts of the Hamiltonian to be effectively "heated". In the approach we will be using for our molecule, the system is split into two hot ($\mathscr{H}$) and cold ($\mathscr{C}$), so that each atom is statically assigned to either the $\mathscr{H}$ or the $\mathscr{C}$ region, and define a parametrized Hamiltonian which depends on λ as follows:

- The charge of the atoms in the $\mathscr{H}$ region is scaled by a factor $\sqrt{\lambda}$.
- The Lennard-Jones parameter ϵ of atoms in the $\mathscr{H}$ region is scaled by a factor λ.
- The proper dihedral potentials for which the first and fourth atoms are in the $\mathscr{H}$ region is scaled by a factor λ.
- The proper dihedral potentials for which either the first or the fourth atom is in the $\mathscr{H}$ region is scaled by a factor $\sqrt{\lambda}$.

This choice of scaling factors allows us to modify only those terms of force-field that are contributing to energy barriers, which are electrostatic, Lennard-Jones and proper dihedrals. This smart scaling results in that:

- interactions inside the $\mathscr{H}$ region are kept at an effective temperature T/λ.

- interactions between the $\mathscr{H}$ and the $\mathscr{C}$ regions are kept at an effective intermediate temperature $T/\sqrt{\lambda}$.
- all interactions inside $\mathscr{C}$ region are kept at temperature T.

For our system all the enhanced sampling was done using HREMD implemented in GROMACS patched with the PLUMED plug-in [14, 41].

2.3 Clustering Algorithm

Cluster analysis is a data analysis tool aimed at grouping elements into categories on the basis of their similarity. There are several strategies about how to cluster the data points. In our work we adopted clustering algorithm discussed in the ref [15].

According to this approach, it has its basis only in the distance between different data points. This algorithm is able to detect nonspherical clusters. The cluster centers are defined as local maxima in the density of data points. It assumes that those centers are surrounded by neighbours with lower local density and that they are at a relatively large distance from any data point with higher local density. For each data point i, we have to compute two quantities: its local density ρ_i and its distance δ_i from the points with higher density. Both these quantities depent only on the distance d_{ij} between data points i and j. When δ_i is much higher than the typical nearest neighbor distance tthan data point i represents local or global maxima in the density. Each of these configurations that correspond to the peak in the density defines a different cluster and all the other configurations are assigned to the cluster whose density peak is the nearest, following the procedure explained in ref 15.

In our case as data points we are choosing the set of slow degrees of freedom (dihedral angles) which can only be fully sampled during the hamiltonian replica exchange molecular dynamics. The number of the chosen dehidral depends on how big and flexible is the molecule, to change conformations. The bigger is the number of the slow degrees the harder becomes to properly clusterize the data. For this reason wise choice of dihedral angles is very important for clustering to be successful.

2.4 Density Functional Theory

the N electron wavefunction is a very complex scalar field, that depends on 3N coordinates. Therefore, it is clear that it is impossible to find even the ground-state wavefunction of biological molecules that have hundreds of electrons. Hence, some simplification is needed. A basic idea to simplify the N electron problem consists in finding some physical quantity that can define the system uniquely without growing in complexity as a function of the number of electrons N. Obviously, this cannot be the N-electron wavefunction. For this purpose Density Functional Theory (DFT) was introduced in the sixties. In DFT the focus is on the electron density rather than on the wavefunction. Since its introduction DFT has become remarkably widespread in the condensed matter physics and quantum chemistry communities because of its good compromise between accuracy and computational cost. The main advantage of the DFT approach is the mapping of a many-body electronic problem onto a noninteracting one. DFT was developed in the work of Hohenberg and Kohn [42] and Kohn and Sham [43].

A practical formulation of DFT that can give an actual recipe of how to solve the quantum many-body problem in practice was provided by Kohn and Sham in 1965 [43]. Kohn and Sham's intuition was to find a one-particle equation within the framework of DFT. These equations are now called the Kohn-Sham (KS) equations. The KS ansatz, which provides us with the expression for the KS energy functional reads in the following way:

$$E_{KS}[n] = T_s[n] + E_H[n] + E_{xc}[n] + \int v_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E_{ion} \tag{2.9}$$

$$n(\mathbf{r}) = 2\sum_{\nu\in occ} |\psi_\nu|^2 \tag{2.10}$$

ψ_ν-s being KS orbitals and $T_s[n]$ being the kinetic energy of a virtual non interacting N electron system with the exact same density as the interacting one;

$$T_s = -\frac{1}{2}\sum_{\nu\in occ} \langle\psi_\nu|\nabla^2|\psi_\nu\rangle \tag{2.11}$$

The Hartree energy is defined as the classical electrostatic energy of a system with electron density $n(\mathbf{r})$:

$$E_H[n] = \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{2.12}$$

and the XC energy E_{xc} has to account for: i) the exchange effects coming from the Pauli repulsion between electrons with the same spin; ii) the correction compensating the self-interaction term in E_H; iii) the correlation effects.

A universal functional of the electron density $F[n]$ is the sum of T_s, E_H and E_{xc}, where the latter (the xc energy) is the difference between what can be computed and what can not.

$$F[n] = T_s[n] + E_H[n] + E_{xc} \tag{2.13}$$

$$E_{xc}[n] = \int_0^{e^2} d\lambda \langle \psi_\lambda | \frac{dV_{int}}{d\lambda} | \psi_\lambda \rangle - E_{Hartree}[n] \tag{2.14}$$

2.4.1 Self-consistency

In practice to find the ground state density an iterative method is used. Initially the starting guess for density $n_0(\mathbf{r})$ is generated. One way to do it is simply to add single atom densities. From the density an effective potential is obtained. With the guessed density and V_{eff} the Kohn-Sham equations can be solved to obtain new set of wavefunctions, from where a new density is computed. If the initial and final densities differ by more than an *a priori* defined small threshold, then the cycle is repeated, but now the new input density is obtained by mixing two previous densities. The better we choose the proportions for mixing two densities, the sooner convergence will be achieved. When the output density is sufficiently close to the input density we assume that convergence has been achieved and the total energy (KS energies), forces and other physical quantities can be computed.

2.5 Time-Dependent Density Functional Theory

DFT is very successful to solve the ground state properties of electrons in time independent potentials, but the main limit of standard DFT is that

it is a ground-state theory, thus not strictly applicable to the calculation of excitation energies or other excited-state properties. Time dependent density functional theory (TDDFT) is the generalization of DFT to the time dependent case. It extends the basic ideas of ground-state density functional theory to the treatment of excitations or more general time-dependent phenomena. As Hohenberg-Kohn theorem is to DFT, Runge and Gross [44] derived a density functional formalism, comparable to the ground state DFT formalism of Hohenberg, Kohn and Sham, for arbitrary time-dependent systems.

Like in DFT theory, the exact density of the many-body system can be obtained from the set of fictitious one particle time-dependent Schrödinger equations:

$$i\frac{\partial}{dt}\Psi_i(\mathbf{r},t) = \Big[-\frac{1}{2}\nabla^2 + v_{KS}(\mathbf{r},t)\Big]\Phi_i(\mathbf{r},t) \tag{2.15}$$

where equations 2.15 are called time dependent Kohn-Sham (TDKS) equations. From the orbitals Ψ_i the density is obtained:

$$n(\mathbf{r},t) = \sum_{i}^{N_v} \Psi_i^*((\mathbf{r},t)\Psi_i((\mathbf{r},t) \tag{2.16}$$

N_v being the number of occupied states.

TDDFT does not give direct access to excited-state properties. However, the knowledge of linear response functions calculated within TDDFT, allows us to estimate energies and transition matrix elements for electronic excitations.

The interaction of a molecule with light can be modeled as the interaction with an electric field varying sinusoidally in time. Excitation energies and oscillator strengths can then be obtained from the poles and residues of the dynamic polarizability, which is defined as:

$$\alpha_{ij}(\omega) \equiv \frac{\partial d_i}{\partial E_j} = -(\hat{r_i}, (\omega - \mathscr{L})^{-1} \cdot [\hat{r_j}, \hat{\rho}^\circ]), \tag{2.17}$$

$\mathbf{d}$ being dipole induced by the perturbing potential, $\hat{r_i}$ is the i-th component of the dipole operator, and the operator $\mathscr{L}$ is defined as:

$$\mathscr{L} = \big[\hat{H}^\circ_{KS}, \rho'(t)\big] + \big[\hat{V'}_{Hxc}[\rho'(t)], \hat{\rho}^\circ\big] \tag{2.18}$$

$\hat{\rho}^\circ$ and $\rho'(t)$ being unperturbed density matrix and the response cause by the time-dependent external potential, respectively, $\hat{H}^\circ_{KS}$ being time independent unperturbed part of hamiltonian and $\hat{V'}_{Hxc}$ being response part, the linear correction to the Hartree plus XC potential.

The right side of the inner product in eq. 2.17 is given by the solution of the linear equation:

$$(\omega - \mathscr{L})\rho = [\hat{r_i}, \hat{\rho^\circ}] \tag{2.19}$$

When the external perturbation is absent, the equation is an eigenvalue equation for the free oscillation modes of the electronic system. The oscillation frequencies, *i.e.* the eigenvalues of this equation, are the excitation energies of the system. This is Casida's approach to TDDFT [45, 46].

$$\begin{aligned} \mathscr{L} v_i &= \omega_i \hat{v}_i \\ \mathscr{L}^\dagger u_i &= \omega_i \hat{u}_i \end{aligned} \tag{2.20}$$

$\hat{v}_i$ and $\hat{u}_i$ being right and left eigenvalues of $\mathscr{L}$. The left eigenvectors can be easily obtained from their corresponding right eigenvectors, therefore, in practice, only one of the equations in 2.20 is solved.

The left and right eigenvectors satisfy the bi-orthogonality condition and identity operator can be constructed from them:

$$\hat{I} = \sum_i |\hat{v}_i\rangle\langle\hat{u}_i| \tag{2.21}$$

Inserting this expression to the one of polarization (eq 2.17) leads to:

$$\alpha_{ij}(\omega) = -\sum_l \frac{(\hat{r_i}, \hat{v}_l)(\hat{u}_l, [\hat{r_j}, \hat{\rho^\circ}])}{\omega - \omega_l + i\epsilon} \tag{2.22}$$

The absorption spectrum from polarization is given as:

$$\kappa(\omega) \propto \frac{1}{3} Im[\alpha_{11}(\omega) + \alpha_{22}(\omega) + \alpha_{33}(\omega)] \tag{2.23}$$

and for each eigenvalue, ω_l, the oscillator strength is given in a following form:

$$F_{ii,l} \equiv -(\hat{r_i}, \hat{v}_l)(\hat{u}_l, [\hat{r_i}, \hat{\rho^\circ}])\pi \tag{2.24}$$

2.6 Ab initio molecular dynamics

2.6.1 Born-Oppenheimer MD

Born Oppenheimer (BO) approximation is introduced by the assumption that the motion of atomic nuclei and electrons in a molecule can be separated. BO approximation relies on the fact that the nuclei are much more massive than the electrons, which allows us to say that the nuclei

are nearly fixed with respect to electron motion. We can fix **R** 3N nuclei coordinates and solve the electronic wavefunction for different values of **R**, where those coordinates of the nuclei play the role of just parameters. If that is done for a range of **R**, the potential energy curve along the nuclei movement is obtained.

In the BO approximation the equations of motion lead to a Schrödinger-like equation for the electrons and a Newton-like equation for the ions (after some assumptions for the ionic wave function) [47]: The Lagrangian for BO dynamics is written in the following way:

$$\mathcal{L}_{BO}(\mathbf{R}, \dot{\mathbf{R}}) = \sum_{I=1}^{N} \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 - \min_{\Phi_i} E^{KS}[\{\Phi_i\}; \mathbf{R}] \tag{2.25}$$

The minimization is constrained to orthogonal sets of Φ orbitals.

The equations of motions for nuclei are

$$M_I \ddot{\mathbf{R}}_I = \nabla_I \left[\min_{\Phi_i} E^{KS}[\{\Phi_i\}; \mathbf{R}] \right] \tag{2.26}$$

During the Born-Oppenheimer MD, DFT energy is calculated at every step of the dynamics, and corresponding forces on the ions are obtained in the following way:

$$\mathbf{F} = -\frac{\partial E}{\partial \mathbf{R}} = -\langle \Psi | \partial \hat{H} / \partial \mathbf{R} | \Psi \rangle \tag{2.27}$$

R nuclei positions are updated according to the Verlet algorithm.

In this thesis all the dynamics for the systems with metal atoms (chapter 4) were run with Born Oppenheimer dynamics.

2.6.2 Car-Parrinello MD

One of the breakthroughs in first principle MD came with the development of Car-Parrinello (CP) MD [16]. CPMD combines the advantages of Ehrenfest MD and BO MD, while avoiding the disadvantages of both of them. The advantage of Ehrenfest MD is that the wave function that minimizes the energy initially, will stay in its respective minimum while the nuclei move. However, the time step is determined by the electronic motion. On the other hand, in BO MD the time step is determined by the slow motion of nuclei, but the electron energy minimization is required at each time step, that is the most time consuming part. Ideally, one would wish to have a time step as large as possible, set by the nuclear motion and at the same time avoid the self consistent energy minimization at every time

step. That is exactly what the CP method offers. During CPMD ionic and electronic degrees of freedom are coupled at each time step, both following classical dynamics. Thus, an electronic minimization is not needed at every time step, after the initial minimization. During fictitious dynamics of the electrons, they will be kept near the ground state as nuclei move. While in BO MD first the electronic structure is optimized and then ions are moved, in CPMD both electrons and ions evolve simultaneously.

The extended Lagrangian can be set up:

$$\mathcal{L}_{CP}(\{\Psi_i\}, \mathbf{R}_n) = \sum_i \frac{\mu_i}{2} \int \Psi_i^* \Psi_i d\mathbf{r} + \sum_{n=1}^{N} \frac{M_n}{2} \dot{\mathbf{R}}_n - E_0(\mathbf{R}, \{\Psi_i\}) + C \quad (2.28)$$

where M_n are the masses of the ions, R_n positions of ions, Ψ_i are the electronic wavefunctions and μ_i are the fictitious masses of electrons with orbital Ψ_i. E_0 is the functional of the orbitals and C can be defined from the orthonormality of the orbitals:

$$C = \sum_{ij} \lambda_{ij} \Big(\int \Psi_i^* \Psi_j d\mathbf{r} - \delta_{is} \Big) \quad (2.29)$$

where λ_{ij} are the Lagrange multipliers ensuring the orthonormality of the Ψ_i orbitals.

If we consider the electronic orbitals as "classical particles", in the extended Lagrangian (eq. 2.28) the first term corresponds to the kinetic energy of electronic orbitals and the second one to the kinetic energy of ions. The third and fourth terms are the potential energy and the constraint (orthonormality condition) respectively.

From the Lagrangian (eq. 2.28) the equations of motion can be derived both for ions and for electronic orbitals:

$$\mu_i \ddot{\Psi}_i = -\frac{\delta E_0}{\delta \Psi_i^*} + \sum_j \lambda_{ij} \Psi_j \quad (2.30)$$

$$M_n \ddot{\mathbf{R}}_n = -\nabla_n E_0(\mathbf{R}, \{\Psi_i\}) + \sum_{ij} \lambda_{ij} \nabla_n \int \Psi_i^* \Psi_j d\mathbf{r} \quad (2.31)$$

The tricky part of CPMD is to correctly choose the fictitious electron masses. The smaller they are the more the electrons follow adiabatically the ionic motion. On the other hand, the time step of the dynamics is chosen according to the vibrational frequencies of the electronic degrees of freedom. The smaller μ_i fictitious masses are the higher are the vibrational frequencies, thus the smaller time step is required to get accurate trajectories.

In practice, choosing the fictitious masses μ we have to compromise between the adiabaticity and computational time. If we select too low fictitious masses, it may happen that CP dynamics would require more computer time than the BO MD, which makes the whole approximation useless. If the masses are set too high, then the electrons won't follow ions adiabatically and the fictitious kinetic energy of electronic orbitals will rise. A wise choice of the fictitious mass μ makes it possible to limit the energy transfer from the nuclear to the electronic degrees of freedom, thus reducing the need of reoptimizing the electronic wavefunctions at each ionic step. In practice to be safe in choice of μ one needs to monitor electronic orbital kinetic energies during the whole dynamics. In a good case it should stay very low compared to the total fictitious energy of the system. Also the mentioned total energy should be conserved during dynamics.

All the results in chapter 3 were obtained by employing CPMD simulations.

2.6.3 QM/MM Dynamics

QM/MM coupling methods are a class of multi-scale schemes in which a quantum mechanical (QM) simulation is embedded in a larger molecular mechanics (MM) simulation. Due to the high computational cost of QM models, this hybrid scheme has become very popular in many scientific disciplines in last few years. The Hamiltonian describing this system is the following:

$$\hat{H} = H^{\hat{MM}}(TOT) - H^{\hat{MM}}(QM) + H^{\hat{QM}}(QM) + H^{Q\hat{M}/MM} \tag{2.32}$$

$H^{\hat{MM}}(TOT)$ is a hamiltonian for whole system calculated with classical molecular mechanics, $H^{\hat{MM}}(QM)$ is a hamiltonian for QM subsystem calculated with classical molecular mechanics, $H^{\hat{QM}}(QM)$ is a hamiltonian part for QM subsystem, calculated quantum mechanically in a DFT framework, $H^{Q\hat{M}/MM}$ describes the hamiltonian for the interaction between the atoms of QM and MM subsystems. $H^{Q\hat{M}/MM}$ contains electrostatic and van der Waals interactions:

$$H^{Q\hat{M}/MM} = \sum_{i \in MM} q_i \int dr \rho(r) v_i(|r - r_i|) + H_{lr} + \sum_{i \in MM, j \in QM} v_{vdw(r_{ij})} \tag{2.33}$$

$$v_i(r) = \frac{r_{ci}^4 - r^4}{r_{ci}^5 - r^5} \tag{2.34}$$

So to sum up, during QMMM simulations only the forces on QM (small) system and the interaction between QM and MM systems are computed quantum mechanically, which saves a lot of computational time compared to BOMD or CPMD simulations.

2.7 Exchange Correlation Functional

There is no approximation in DFT, so the problem is formulated in an exact way, but we do not know any functional form for the exchange-correlation term – that is when we are forced to find an approximation to the many-body problem. Even though the exact functional for xc energy functional must be very complex, various simple approximations exist that lead to quite accurate results for various systems. The functionals used in this thesis are Generalized Gradient Approximations (GGA) and Hybrid Functionals.

- In the GGA the xc energy functional depends on the density n and on its gradient $|\nabla n|$. In this way the changes in the density of the system are accounted for by including the gradient of the density in the xc energy functional unlike the simplest Local Density Approximation (LDA), where the exchange functional depends only on the density of the system at each position. The generalized form of the GGA functional can be written in the following way:

$$\begin{aligned} E_{xc}^{GGA}[n,n] &= \int d^3r n(\mathbf{r})\epsilon_{xc}(n,|,\nabla_n|,...) \\ &\equiv \int d^3r n(\mathbf{r})\epsilon_{xc}^{hom}(n)F_{xc}(n,|\nabla_n|,...) \end{aligned} \tag{2.35}$$

  Where ϵ_{xc}^{hom} is the exchange energy of the unpolarized gas and for F_{xc} several expressions are available.

- The hybrid functionals are called "hybrid" because they are a combination of orbital-dependent Hartree-Fock and an explicit density functional. So far they are the most accurate functionals available as far as energetics is concerned and are widely used in the quantum chemistry community.

$$E_{hybr} = \alpha E_{HF}^{x} + (1-\alpha)E_{GGA}^{x} + E^{c} \tag{2.36}$$

The most striking change due to use of hybrid functionals is the predicted excitation energies. In equation 2.14, when $\alpha = 0$ the exchange correlation energy is just the Hartree-Fock term and when $\alpha = 1$ (full coupling)

LDA or GGA functional is the most appropriate. Later Becke introduced parametrized forms that are accurate for many molecules and were used for the calculations in this thesis. The new xc functional consists in a three-parameter functional, mixing HF exchange, exchange functional of Becke (B88) and LYP correlation with different weights obtained empirically to fit atomic and molecular data.

For all the ab initio molecular dynamics simulations we have used GGA functional and for absorption spectrum capculations B3LYP [48] hybrid functional was used.

Chapter 3

Molecular mechanisms of color expression in anthocyanins

3.1 Cyanin

As we have mentioned in section 1, hundreds of different anthocyanins have been isolated from plants. While they are all based on a single basic core structure – the flavyllium ion (see 1.2) the side groups attached to the chromophore may differ in each structure. These side groups can be a hydrogen atom, a hydroxide or a methoxy-group. In addition, diversity is further increased by the chemical combination of sugars with organic acids (such as acetic, succinic, caffeic acid and many more) to produce acylated anthocyanins. Up to now 23 anthocyanidins (see chapter 1) have been discovered, from where the most common species present in plants are Cyanidin, Pelargonidin, Delphinidin, Peonidin, Petunidin, Malvidin (see figure 1.4). The distribution of anthocyanins in edible parts of plants may be summarized as follows: Anthocyanins based on cyanidin occur most frequently. The percentage occurrence of the anthocyanidins can be estimated approximately: cyanidin (50%); pelargonidin, peonidin, and delphinidin (each 12%); and petunidin and malvidin (each 7%) [49]. Thus, the cyanidin type predominates over the other anthocyanidins. In terms of glycoside distribution, the 3-monoside is the most primitive glycosidic pattern. Moreover, 3-glycosides occur about two and a half times more frequently than 3,5-diglycosides. Summing up all these statistics the most ubiquitous anthocyanin happens to be cyanidin-3-glucoside. That is why we chose cyanidin-3-glucoside (from now referred as C3G) as the main focus of our study.

Energy differences (kcal/mol)			
	Pelargonin	**Cyanin (C3G)**	**Delphinin**
$A^{\circ}_{4'} - A^{\circ}_{7}$	**0.7**	**-15**	**-3**
$A^{\circ}_{4'} - A^{\circ}_{5}$	**0.8**	**-26**	**-3**
$A^{-}_{4'7} - A^{-}_{4'5}$	**0.6**	**-5**	**0.5**
$A^{-}_{4'7} - A^{-}_{57}$	**-4**	**-24**	**-8**

Table 3.1: The relative energy (kcal/mol) of different protomers of different ANT secies investigated

Placing C3G in aqueous solutions with pH values ranging from strongly acidic pH=1 to neutral base pH=9 results in the production of very different colors. On the basis of such visual observations, it is clear that a great number of elementary reactions must be involved. As we have already mentioned in section 1 there are quite a few stable forms of anthocyanins depending on the pH of the solution. All the previous computational studies have focused on the cationic form [6, 7, 9, 10, 17, 50, 51], basically because full protonation lifts any ambiguity as to its tautomeric state and makes it the only relevant species at low pH, thus, facilitating the comparison with experiments [1, 3–5, 52]. In addition, in this work we also investigate neutral and negative species. We performed geometry optimizations at DFT-B3LYP [48] level of theory with the `Gaussian09` [53] programs to establish relative energies of different tautomers of the neutral and negative anthocyanins and for two other members of the C3G family (pelargonidin and delphinidin) in presence of implicit solvent [64]. The B3LYP [48], exchange correlation functional was used. Finally, for further investigation two out of the three protomers of both the neutral and negative species of C3G were selected according to their relative DFT energies (see Table 3.1). From these energies we exclude protomers A°_{5} and A^{-}_{57} in the case of C3G. We thus expect C3G in solution to be predominantly a mix of the remaining five charge/protomeric states, whose relative abundance depend on the

pH. In this chapter we aim at disentangling the structural, dynamical, and electronic features responsible for the color optical properties of C3G. For that we chose main stable forms of C3G that co-exist at various pH conditions. These forms are posively charged flavyllium ion referred as A^+ from now on, two tautomeric species of neutrally charged C3G referred as $A^\circ_{4'}$ and A°_7, and two negative tautomeric species $A^-_{4'7}$ and $A^-_{4'5}$ (See figure 3.1).

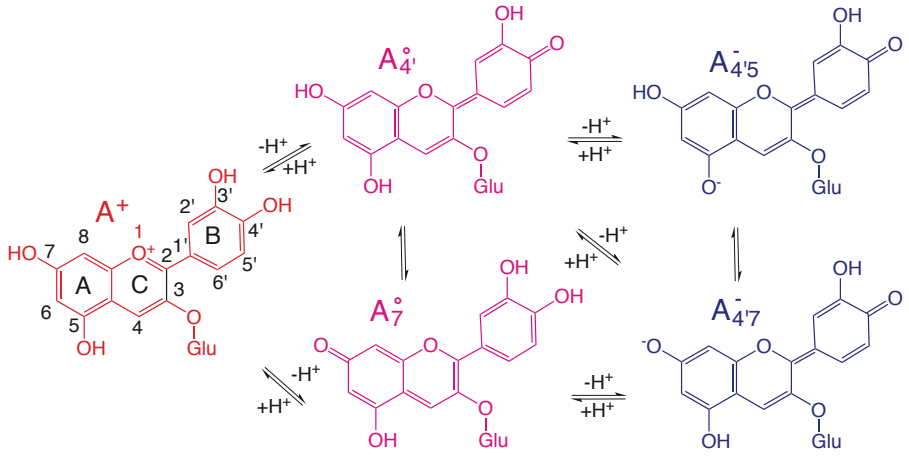


Figure 3.1: Tautomeric forms of cyanidin-3-glucoside in solution at pH 1-9.

To this end, we show the results obtained by applying the multi-scale computational protocol introduced in previous chapter to C3G molecule to study its optical properties.

3.2 Conformational Landscape

Classical MD simulations were run using `GROMACS` 4.5.5, [41] with all bonds to hydrogen atoms constrained using LINCS [54], upon adequate equilibration. General AMBER Force Field (GAFF) [21] parameters were assigned using the `antechamber` module of `AmberTools13`, with RESP charges [55], at the HF/6-31G* level, calculated on DFT optimized geometries. All above mentioned species of C3G were simulated at 300K for 1 μs fully solvated in TIP3P water [56], using cubic cells of 30 Å edge accommodating $\sim$ 1000 water molecules, in periodic boundary conditions. The volume was kept fixed, and velocity rescaling was used to keep the average temperature constant (NVT ensemble). In order to enhance the

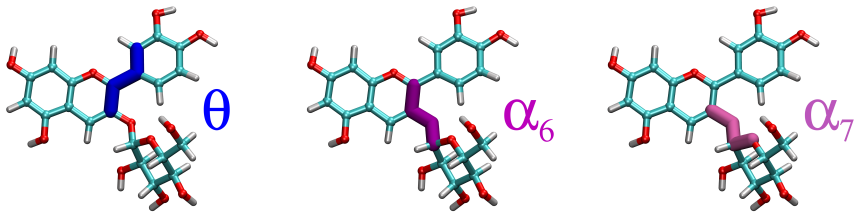


Figure 3.2: θ, α_6 and α_7 dihedral angles are shown in blue, dark and light magenta, respectively, in the specific case of the favylium cation.

ergodicity of the conformational sampling, Hamiltonian replica-exchange MD (HREMD) [14] was used, as implemented in the `PLUMED` plugin [57] with ten different replicas for each simulation. To ensure ergodicity and convergence the dihedrals of different replicas were checked to match each other for each simulated species.

For each molecular state, our long HREMD trajectory was analyzed in terms of the three dihedrals θ, α_6 and α_7, (Fig. 3.2) in order to find representative geometries corresponding to conformers separated by free energy barrier of the order of a few k_BT, impossible to overcome in AIMD trajectories. These dihedrals were selected as they describe distortions of the chromophore, which would in turn affect the color. First angle θ determines the orientation between the aromatic AC and B rings (Fig. 3.1) and two other dihedrals (α_6 and α_7) are responsible for the sugar orientation in C3G. An advanced clustering algorithm [15] (see 2.3) was used to identify the most representative conformers, according to the most important degrees of freedom (three dihedrals shown in fig. 3.2) of cyanin.

According to the clustering analysis conformers are detected as high density clusters in the dihedral space 2.3. Figure 3.3 shows distributions of each dihedral angle for each investigated species. When the most probable configurations (the centres of clusters) are identified, all the rest of the configurations are distributed into those clusters. Usually several clusters are identified for each species of C3G, and the configurations belonging to each cluster (conformer) are separated to the configurations from the other clusters by high enough free energy barriers, which even during long AIMD simulation the system is not able to overcome. Therefore the free energy landscape is not know quantum mechanically, thus the statistical

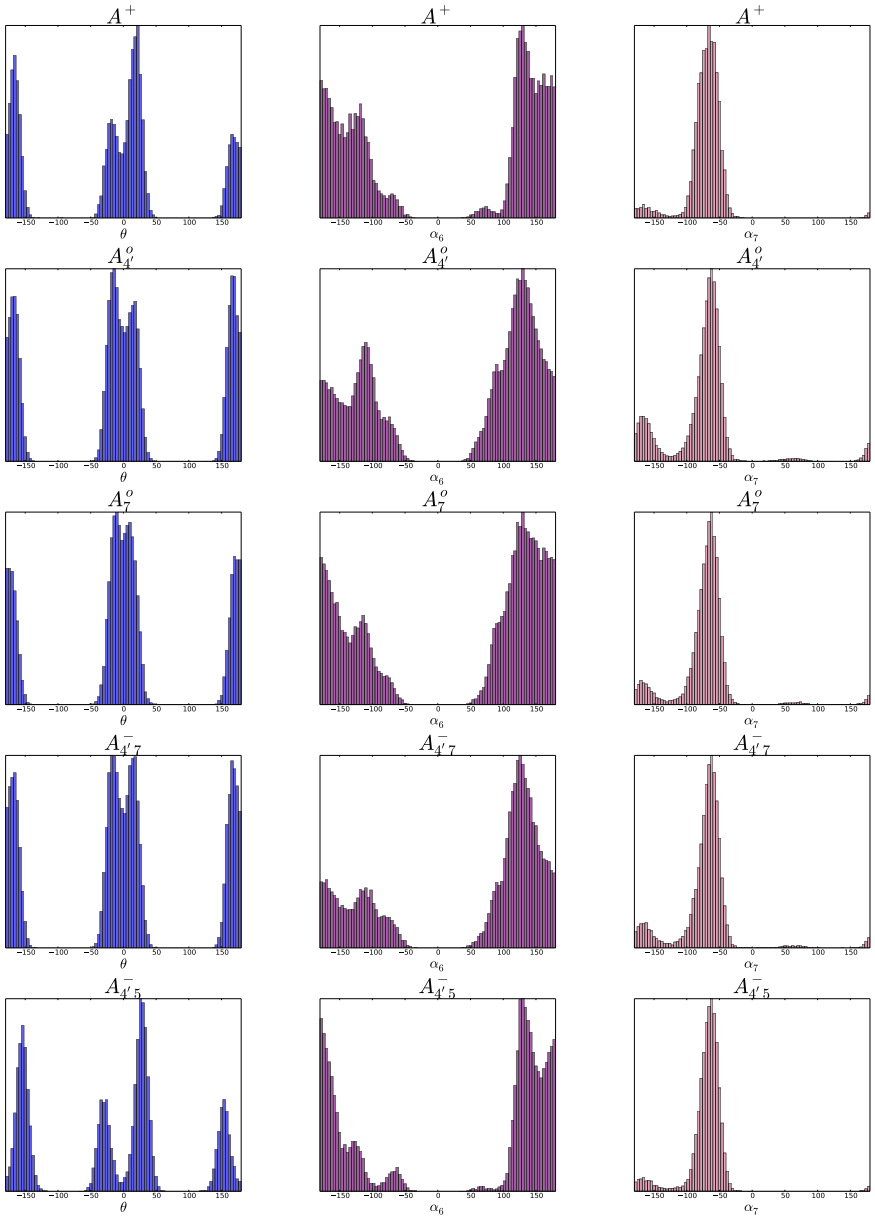


Figure 3.3: θ, α_6 and α_7 dihedral angle distributions from HREMD for all the studied species: A^+, $\mathsf{A}^\circ_{4'}$, A°_7, $\mathsf{A}^-_{4'7}$, $\mathsf{A}^-_{4'5}$.

weight of each cluster is evaluated as the normalized ratio between the residence time within each free-energy basin and the total simulation time of classical HREMD. Since the probabilities of the clusters is computed from data coming from HREMD it is crucial to make sure that the choice of force field is adequate.

For each species we identify the clusters whose probabilities are more than 10%. In all species of C3G we have four relevant clusters, that correspond to four orientations of the θ dihedral angle. On figure 3.5 we

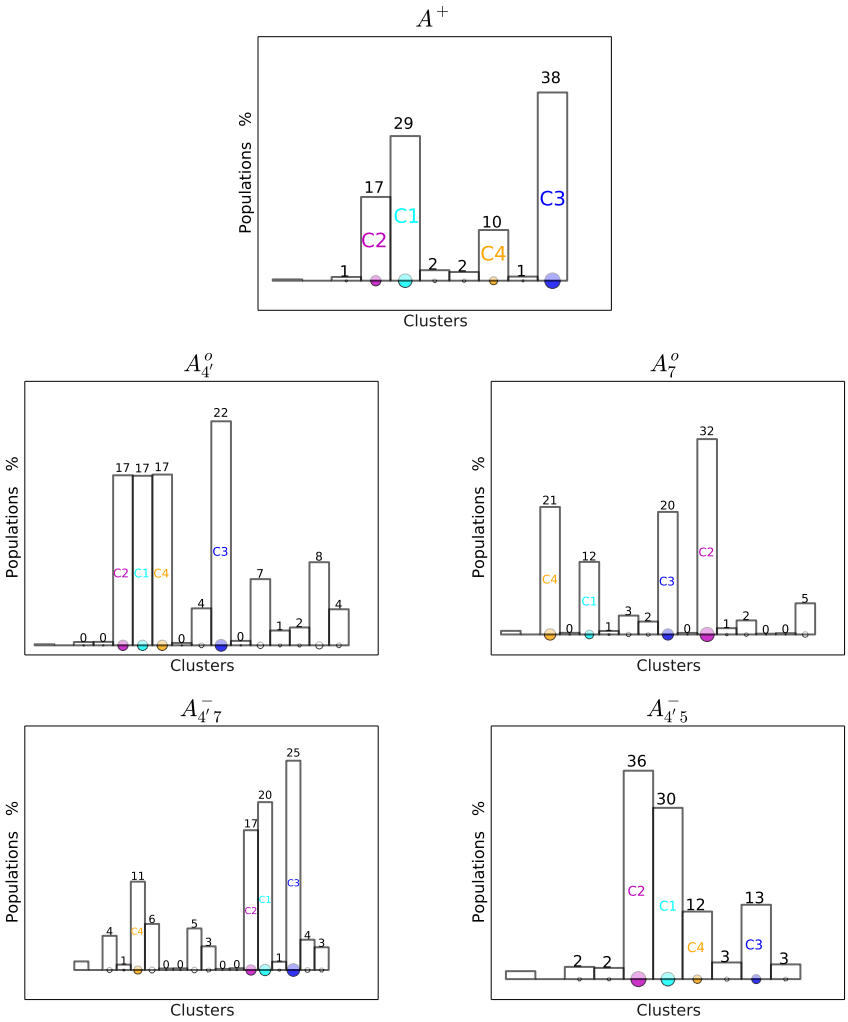


Figure 3.4: The relative populations of clusters as obtained by the analysis of HREMD trajectory of the A^+, $\mathsf{A}^\circ_{4'}$, A°_7, $\mathsf{A}^-_{4'7}$ and $\mathsf{A}^-_{4'5}$ tautomers, shown from top left to bottom right.

show the pair-wise distributions, depicted as 2D cuts of a 3D density map,

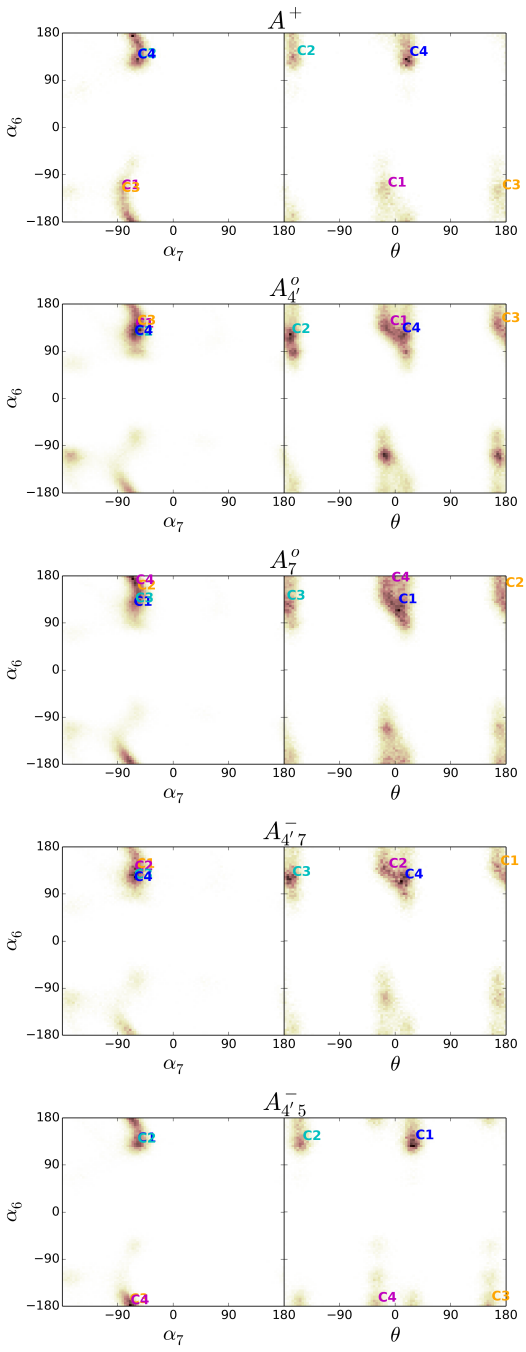


Figure 3.5: 2D density maps of pairs of dihedral angles (in degrees) resulting from the HREMD trajectory for all possible states (A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$ and $A^-_{4'5}$) of C3G and used for the cluster analysis performed to select the dominant conformers.

of three dihedral angles for all investigated tautomers. From this analysis we found that θ is correlated to the position of the sugar, and that each species may adopt four main conformations (each populated for more than 10% of the whole simulation time), corresponding to $\theta \approx 180^{\circ+}$ (C_1), $0^{\circ-}$ (C_2), $0^{\circ+}$ (C_3), $180^{\circ-}$ (C_4) (Fig. 5.3).

Clustering analyses reveals that:

- For the flavylium cation state (A^+) C3 conformer is the most populated (38%), followed by the C1 (29% percent), C2 (17%) and C4 (10%), (Figure 3.4).
- $\mathsf{A}^{\circ}_{4'}$ tautomer of neutral quinonoid state: C1 (17% percent), C2 (17%), C3 (22%),and C4 (17%), (Figure 3.4).
- A°_{7} tautomer of neutral quinonoid state: C1 (12% percent), C2 (32%), C3 (20%),and C4 (21%), (Figure 3.4).
- $\mathsf{A}^{-}_{4'7}$ tautomer of negative quinonoid state: C1 (20% percent), C2 (17%), C3 (25%),and C4 (11%), (Figure 3.4).
- $\mathsf{A}^{-}_{4'5}$ tautomer of negative quinonoid state: C1 (30% percent), C2 (36%), C3 (13%),and C4 (12%), (Figure 3.4).

All of these conformers were chosen for AIMD simulations to obtain more accurate geometries for spectrum calculations.

All the previous computational studies were done on a reduced system, dismissing sugar, as it does not directly participate in the optical transitions on the conjugated core. It has nevertheless a major effect on the geometry of the chromophore, making it of importance when considering the color properties. The relative orientations of the AC and B rings always deviate from planarity, due to steric hindrance of the substituents on the aromatic rings [58], while the varying position of the sugar results from competition between steric and solvent effects.

3.2.1 High frequency dynamics and optical spectra

In order to accurately account for thermal fluctuations, explicit-solvent Car-Parrinello AIMD simulations [16, 59] were performed for each one of the four most populated conformers of the five C3G species investigated in this work, (A^+, $\mathsf{A}^{\circ}_{4'}$, A°_{7}, $\mathsf{A}^{-}_{4'7}$, and $\mathsf{A}^{-}_{4'7}$), using the `cp.x` module of the QUANTUM ESPRESSO distribution. [60, 61] Ultrasoft pseudopotentials

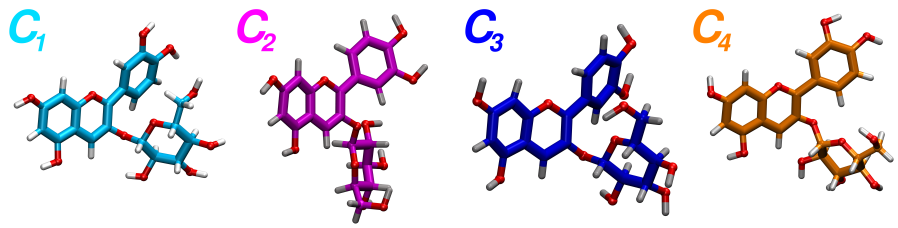


Figure 3.6: Geometries of the four most representative conformers of the flavylium cation. Qualititatively similar results hold for the other molecular states investigated here.

Length of CP trajectories (ps)					
Conformer Name	A^+	$A^\circ_{4'}$	A°_7	$A^-_{4'7}$	$A^-_{4'5}$
C1	**15**	**40**	**30**	**20**	**20**
C2	**15**	**25**	**15**	**15**	**20**
C3	**15**	**40**	**25**	**20**	**20**
C4	**15**	**20**	**15**	**15**	**20**

Table 3.2: The length of CP trajectories for each anthocyanin species investigated A^+, A_4, A_7, $A^-_{4,7}$ and $A^-_{4,5}$.

from the QUANTUM ESPRESSO public repository were used the PBE exchange correlation-functional [62] with plane-wave basis sets with cutoffs of 25 and 200 Ry for wavefunctions and density, respectively. The number of water molecules was reduced to ~ 200, so as to fit a simulation cell of linear dimensions ~ 20 Å. The masses of the hydrogen and oxygen atoms were scaled to be equal to the physical mass of carbon atoms (12 a.u.), which optimizes autocorrelation times, while hindering thermalization of the fictitious electronic degrees of freedom, MD time step of 5 a.u. ($\sim$0.12fs). The simulation box including explicit solvent was pre-equilibrated via classical MD at constant pressure/temperature. AIMD simulations were finally performed at constant volume and (room) temperature using the Nosè-Hoover thermostat [63] for at least 15 ps per equilibrated trajectory. See Table 3.2 for the lengths of CP trajectories for each examined species.

As a check, we make sure that the dihedral distributions coming from

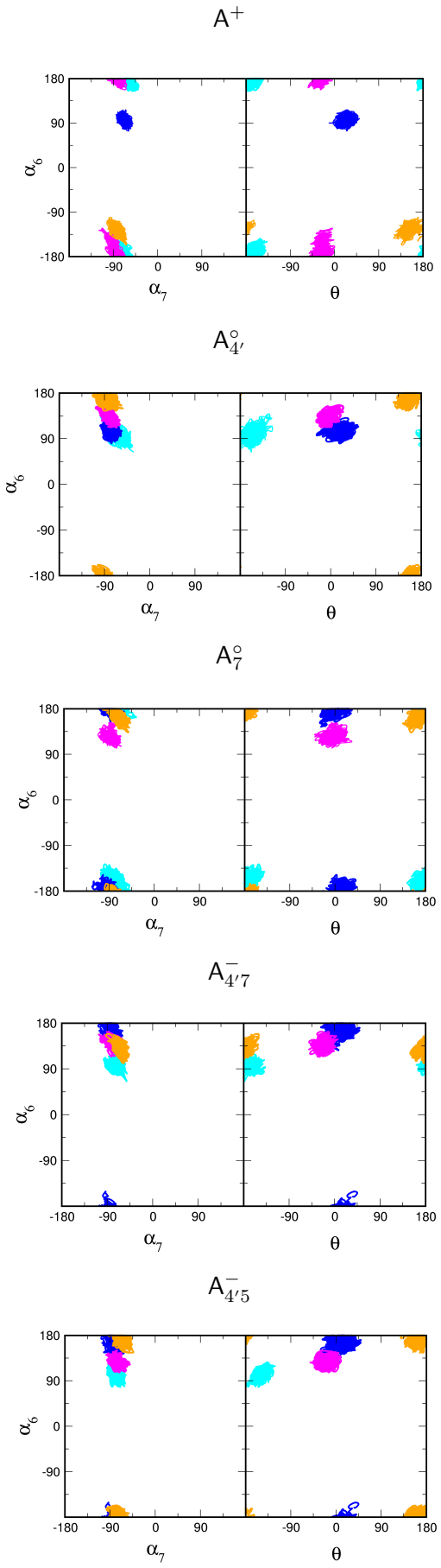


Figure 3.7: α_6 vs α_7, α_6 vs θ in AIMD calculations (color code corresponds to conformers:C1, C2, C3, C4 .

the AIMD trajectories match the ones from the classical simulations. This is done by comparing the heatmaps presented in figure 3.5 with the data presented in figure 3.7. The better the agreement between them, the more we are sure that the force field used for the HREMD was good enough and that the ab-initio simulations are ergodic enough. In general, AIMD well reproduces the dihedral values obtained from classical MD, with small rearrangements of sugar position at times.

While the GGA exchange correlation functional was accurate enough for sampling AIMD trajectories, its reliability becomes questionable for spectrum calculations. It is known that GGA XC functionals tend to underestimate low-lying transition energies. For these reasons we have chosen to use more accurate XC functionals, namely the hybrid B3LYP functional. The explicit size of our system (solute + ∼200 explicit water molecules) would make B3LYP calculations computationally extremely expensive. Thus, we have adapted the *self-consistent continuum solvation* (SCCS) approach [17] based on TDDFT, implemented in QUANTUM ESPRESSO. According to this method spectrum calculations are done on the solute in implicit solvent, while AIMD calculations were run in an explicit solvent. In this way both thermal fluctuations and dielectric solvation effects on optical spectra are accounted.

Initially, analysis of the H-bond persistence was done with the ptraj module of Ambertools 13 [21]. The absorption spectrum calculations were performed on molecular configurations selected from the AIMD trajectories at every 0.5 ps, ensuring statistical independence. The B3LYP [48], exchange correlation functional was used and 50/200Ry wavefunction cut offs were selected to compute the absorption spectra. The solvation model was enhanced by explicitly keeping solvent molecules which were found (from out H-bond persistence analysis) to be persistently H-bonded to the chromophore (typically, at positions 4′, 3′, 5 and 7: see Table 3.4) and replacing the rest of explicit water molecules with an effective medium [64], in the spirit of a continuum solvation model [65], (see Fig: 3.8), thus accounting for the protic character of the solvent on the electronic structure of the solute. Finally, the spectrum of a given AIMD trajectory was computed as the average of the spectrum of each selected frame. This protocol was applied to all relevant conformers (C1, C2, C3, C4) of each investigated C3G tautomeric states (A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$, $A^-_{4'5}$) and the average spectrum for each of the mentioned case is shown on figure 3.10. In the figure it is clear that there is huge difference not only between different species of

		C1	C2	C3	C4
$\mathbf{A}^{+}$	θ	$-164^\circ \pm 25^\circ$	$-25^\circ \pm 25^\circ$	$25^\circ \pm 25^\circ$	$152^\circ \pm 30^\circ$
	α_6	$-175^\circ \pm 30^\circ$	$-178^\circ \pm 25^\circ$	$95^\circ \pm 20^\circ$	$-125^\circ \pm 25^\circ$
	α_7	$-65^\circ \pm 25^\circ$	$-85^\circ \pm 20^\circ$	$-68^\circ \pm 17^\circ$	$-80^\circ \pm 20^\circ$
$\mathbf{A}^{\circ}_{4'}$	θ	$-153^\circ \pm 25^\circ$	$-10^\circ \pm 25^\circ$	$18^\circ \pm 30^\circ$	$157^\circ \pm 25^\circ$
	α_6	$100^\circ \pm 30^\circ$	$130^\circ \pm 20^\circ$	$105^\circ \pm 20^\circ$	$170^\circ \pm 25^\circ$
	α_7	$-65^\circ \pm 25^\circ$	$-80^\circ \pm 20^\circ$	$-75^\circ \pm 15^\circ$	$-85^\circ \pm 25^\circ$
$\mathbf{A}^{\circ}_{7}$	θ	$-162^\circ \pm 25^\circ$	$-6^\circ \pm 30^\circ$	$7^\circ \pm 25^\circ$	$170^\circ \pm 30^\circ$
	α_6	$104^\circ \pm 20^\circ$	$175^\circ \pm 30^\circ$	$170^\circ \pm 30^\circ$	$165^\circ \pm 25^\circ$
	α_7	$-65^\circ \pm 25^\circ$	$-85^\circ \pm 30^\circ$	$-87^\circ \pm 25^\circ$	$-70^\circ \pm 20^\circ$
$\mathbf{A}^{-}_{4'7}$	θ	$-167^\circ \pm 25^\circ$	$-21^\circ \pm 25^\circ$	$11^\circ \pm 36^\circ$	$177^\circ \pm 27^\circ$
	α_6	$97^\circ \pm 20^\circ$	$-140^\circ \pm 25^\circ$	$165^\circ \pm 20^\circ$	$-130^\circ \pm 25^\circ$
	α_7	$-75^\circ \pm 20^\circ$	$-77^\circ \pm 23^\circ$	$-80^\circ \pm 22^\circ$	$-65^\circ \pm 15^\circ$
$\mathbf{A}^{-}_{4'5}$	θ	$-146^\circ \pm 30^\circ$	$-25^\circ \pm 25^\circ$	$27^\circ \pm 25^\circ$	$166^\circ \pm 26^\circ$
	α_6	$-103^\circ \pm 22^\circ$	$125^\circ \pm 17^\circ$	$107^\circ \pm 20^\circ$	$165^\circ \pm 32^\circ$
	α_7	$-77^\circ \pm 15^\circ$	$-75^\circ \pm 17^\circ$	$-53^\circ \pm 20^\circ$	$-70^\circ \pm 25^\circ$

Table 3.3: Most relevant dihedrals (same used for clustering algorithm) for all the protomers: A^{+}, $\mathsf{A}^{\circ}_{4'}$, A°_{7}, $\mathsf{A}^{-}_{4'7}$ and $\mathsf{A}^{-}_{4'5}$. Means and standard deviations reported. The data is coming from ab-initio MD simulations.

C3G, but also among different conformers of a given species. It further confirms the importance of the full conformational study of these molecules. The peak positions of the absorption spectrum of all studied conformers (C1, C2, C3, C4) of each investigated C3G tautomeric states (A^+, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$, $A^{-}_{4'5}$) are shown in table 3.5.

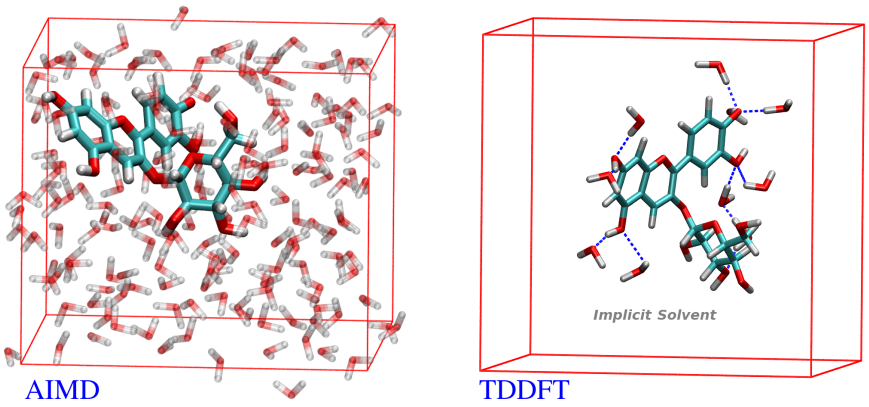


Figure 3.8: On the left: The box of AIMD simulations, C3G molecule in explicit water molecules; On the Right: The box of TDDFT simulations, C3G molecule with few explicit water molecules (H-bonded with chromophore) and continuum solvent [64].

3.2.2 Absorption spectra

The overall spectrum for the solution of each species (A^+, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$, $A^{-}_{4'5}$) was computed by averaging the TDDFT spectra of each conformer (C1, C2, C3, C4), weighted by the factors (see figure 3.4) computed in the conformational analysis . The resulting color was finally estimated using the tri-stimulus model [67]. The average spectra with its corresponding color for each species are shown on the figure 3.10.

As we know the only stable species at highly acidic conditions is the flavylium cation, thus the overall spectrum of A^+ that has its maximum at 512nm is remarkably close to the experimental spectra at ph 1-2, which have their maxima at 509 and 513nm, respectively. The spectra for $A^{\circ}_{4'}$ and A°_{7} obtained as weighted average have their maxima at 534nm and 531nm, respectively, again being in a good agreement with the experimental spectra at pH = 4-6, which have their maxima at 523-528 nm (Figure 7), considering the fact that we are neglecting the contribution of other species

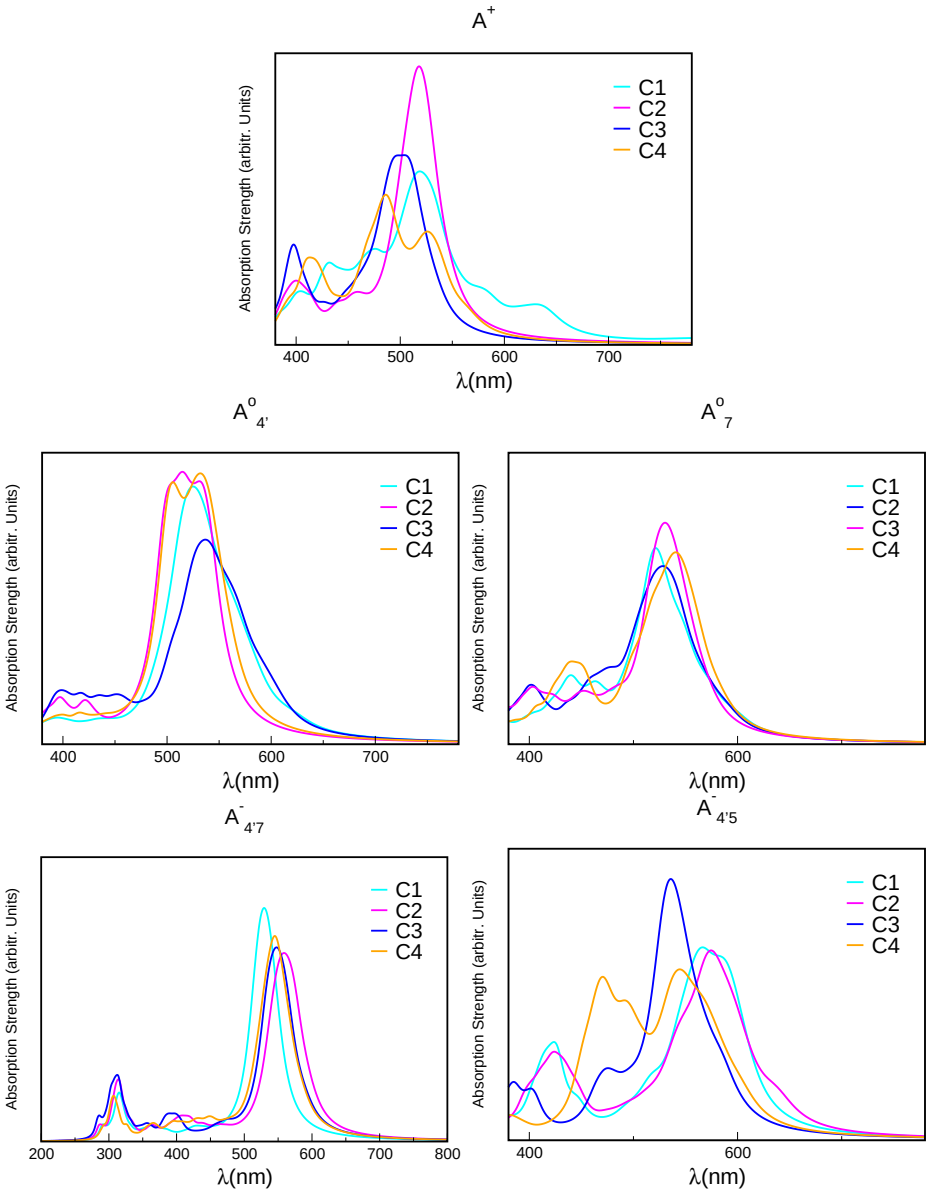


Figure 3.9: The average absorption spectrum along the AIMD trajectory for conformers C1, C2, C3, C4 of each investigated C3G tautomeric states (A^+, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$, $A^{-}_{4'5}$).

Number of H-bonded waters					
Conformer Name	A^+	A_4	A_7	$A^-_{4,7}$	$A^-_{4,5}$
C1	7	6	7	7	7
C2	7	7	6	8	7
C3	8	6	6	7	6
C4	7	6	6	8	8

Table 3.4: Number of water molecules persistently H-bonding to each conformer of all anthocyanin species investigated as obtained from H-bonds analysis of the ab initio trajectories. Analysis of the H-bond persistence was done with the `ptraj` module of Ambertools 13. We remark that H-bond persistence may be overestimated due to the scaled massed used in the AIMD simulations [66].

λ_{max} (nm)					
Conformer Name	A^+	A_4	A_7	$A^-_{4,7}$	$A^-_{4,5}$
C1	519	524	520	529	583
C2	517	529	530	560	574
C3	502	536	528	547	538
C4	527	531	541	546	546

Table 3.5: The position of the absorption spectra peak.

present in solution at this pH range such as A^+ which would shift the λ_{max} of the spectrum to the lower wavelength, making it more comparable to the experiment.

Finally, at pH ranges 7-9 the main contribution in the spectrum comes from the negative species. Spectoscopy results show that the first excitation is peaked around 555-568 nm. The spectrum obtained as a weighted average of the different conformers for $\mathsf{A}^-_{4'7}$ and $\mathsf{A}^-_{4'5}$ has its maximum at 550-545 nm. Also, the $\mathsf{A}^-_{4'5}$ species has the second higher energy absorption peak centred around 470 nm while according to the experimental results the second peak is at 450 nm.

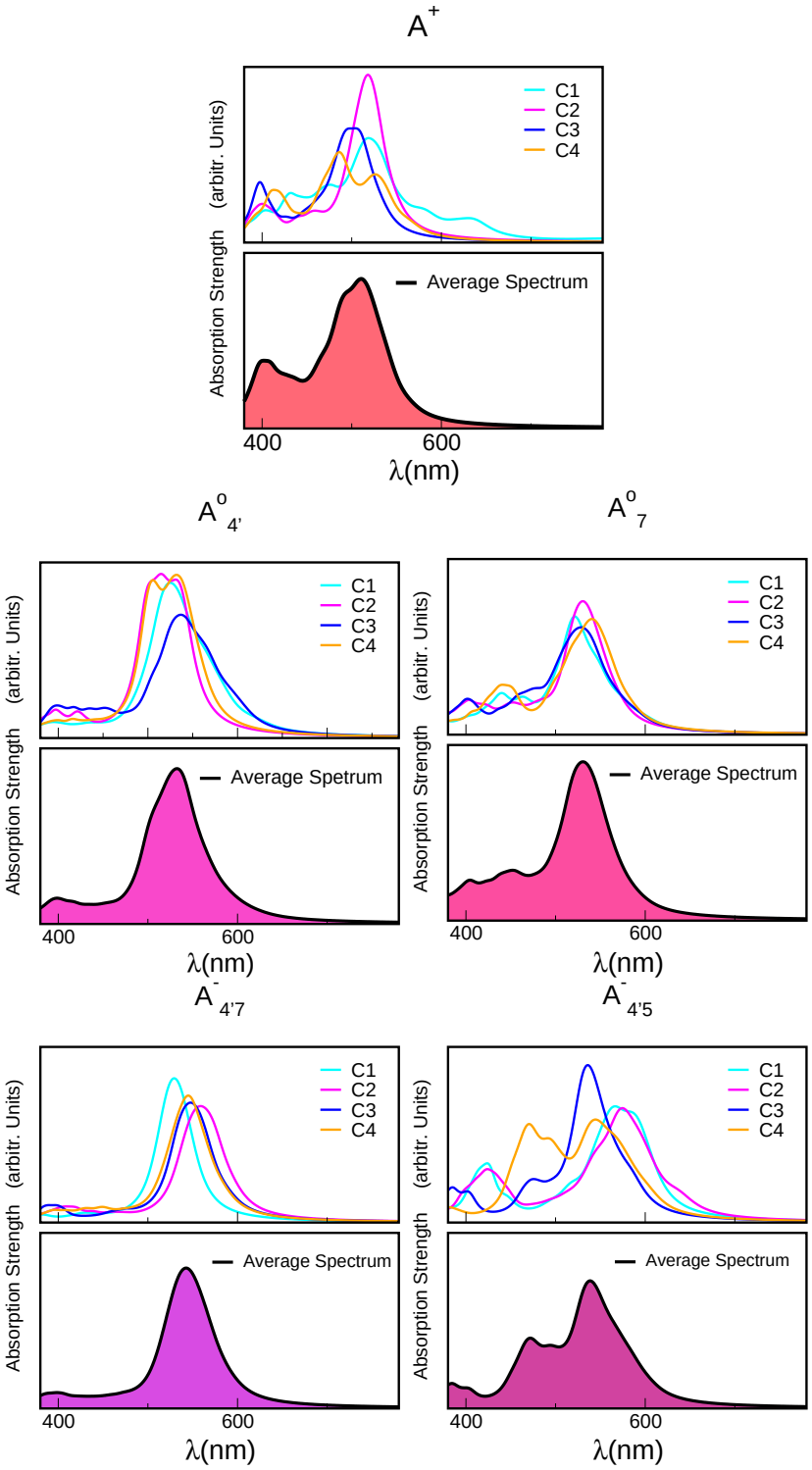


Figure 3.10: Upper Pannels: The average absorption spectrum along the AIMD trajectory for conformers C1, C2, C3, C4 of each investigated C3G tautomeric states (A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$, $A^-_{4'5}$). Down Pannels: Weighted average spectrum for each species and the filling color shows the color each species have in solution.

3.3 Main features affecting optical spectra

3.3.1 pH, charge, and tautomeric state

Experimentally measured and computed spectra are displayed and compared with each other in Fig. 3.11. Experimental pH ranges are matched to the charge state that is prevalent in each of them. Our predictions are in excellent agreement with experiment, as concerns the position and pH/charge dependence of the maximum absorption wavelength, λ_{max}. The latter is determined by the $S_0 \rightarrow S_1$ transition, which is essentially a single-configuration HOMO $\rightarrow$ LUMO transition ($\pi \rightarrow \pi*$, accounting for $\gtrsim 90\%$ of the total oscillator strength). We also note the existence of a shoulder in the spectra predicted for the neutral and negative species at shorter wevelengths ($\lambda \lesssim 500$ nm) and corresponding mostly to a HOMO-1 $\rightarrow$ LUMO transition. This shoulder develops into a full-blown secondary peak in $\mathsf{A}^-_{4'5}$, which is remarkably visible in experimental spectra from intermediate to high acidity. This indicates that $\mathsf{A}^-_{4'5}$ is present in solution at room temperature in this pH range, contrary to the higher stability of $\mathsf{A}^-_{4'7}$ predicted by our static computations (see Table 3.1). Although this discrepancy could be due to the inadequacy of the energy functional or PCM model utilized here, yet unpublished study [68] in our group on the free-energy difference between tautomers of C3G indicate that these may depend largely on entropic effects once thermal fluctuations and the molecular nature of the solvent are explicitly accounted for. We notice that this secondary peak has an outstanding effect on the color expressed by the solution, as it subtracts blue components from the light diffused by the solution. We conclude that hindering deprotonation from site 5, *e.g.* by glycosilation, will likely enhance the *blueness* of the solution.

HOMO-LUMO (frontier orbitals) energies and shape were obtained via single point calculations with localized basis set package gaussian [53] with B3LYP functional and basis set 6-311g(d,p) in order to obtain reliable relative energies. The positions of the relevant frontier energy levels (HOMO-1, HOMO, and LUMO) are shown in Fig. 3.12, along with a representation of the corresponding molecular orbitals, for the five molecular states of C3G. These data show that deprotonation determines a lift in the HOMO and LUMO orbital energies, in line with chemical intuition that makes one expect a decrease of the ionization potential (*i.e.* increase of HOMO energy), when increasing the negative charge of the molecule. To a smaller extent a similar argument also holds for the LUMO, thus leading to a progressive

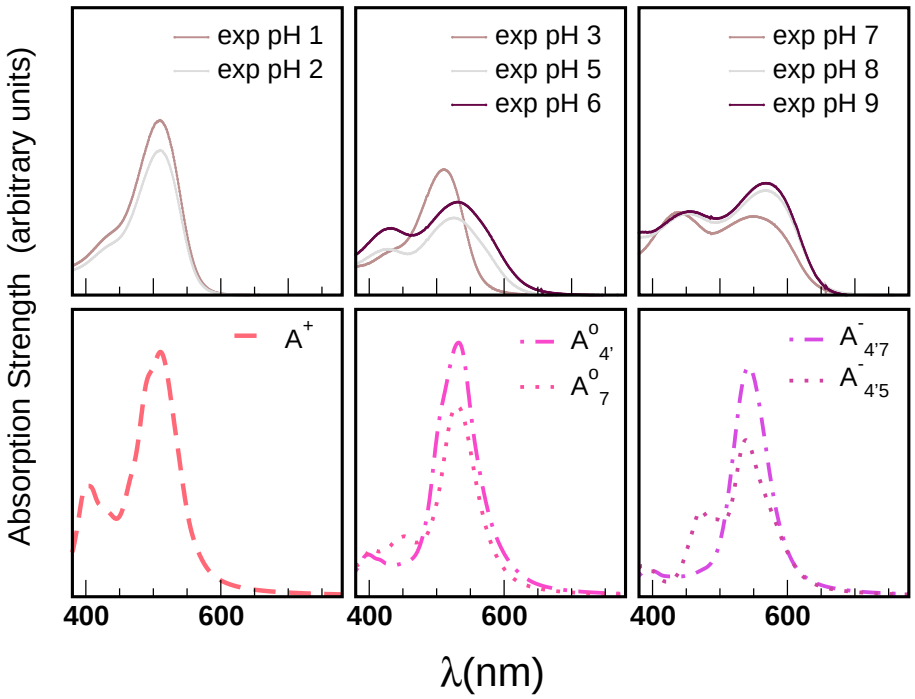


Figure 3.11: Experimental spectra at different pH values (top) and calculated for the species expected to be relevant for each pH interval.

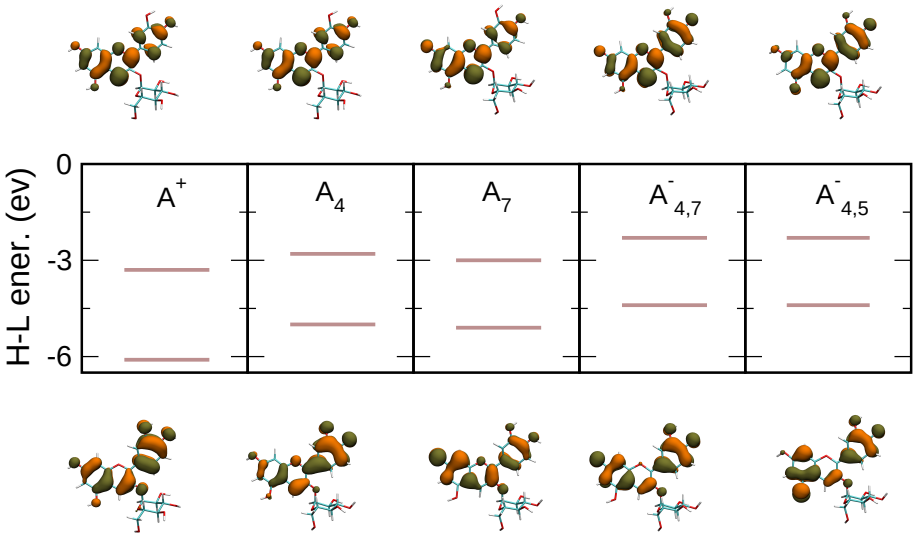


Figure 3.12: HOMO-1, HOMO and LUMO energy levels for all five molecular states, along with the frontier orbitals. The HOMO-LUMO gap is the spacing between the light brown lines while the HOMO-1 level is drawn as a darker line.

decrease of the HOMO-LOMO (HL) gap from the positive to the negative states of C3G. Hence, the charge state of the molecule itself markedly affects the cyanin optical properties, consistently with experimental observations. Both the HOMO and the LUMO lie in the plane of the chromophore and extend over both conjugated moieties (AC and B rings); hence we also expect that distortions in the chromophore, which decrease its overall conjugation, will result in a shift of the orbital energies, thus affecting the color of the solution.

3.3.2 Conformational effects on absorption spectra

The molecular distortion that mostly affects the optical properties of anthocyanins is the twist between the single (B) and double (AC) rings, which is described by the dihedral angle θ. Due to the asymmetry introduced by the sugar and the hydroxyl groups of the B ring, of each conformer C1, C2, C3, C4 of each investigated C3G tautomeric state (A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$, $A^-_{4'5}$) is distinctly deviating from planarity (see figure 3.15 and table 3.3). Moreover, steric hyndrance exists between H1′ (C1 and C3 conformers) H6′ (C2 and C4 conformers) of the B ring and O1′, O2′

O5′ of the sugar (see figure 3.13 and table 3.6), contributing to the relative internal twist of the AC and B rings and out of plane bending of the B ring. The extent of the fluctuations along θ are tightly determined by the way the B ring is trapped with respect to the sugar.

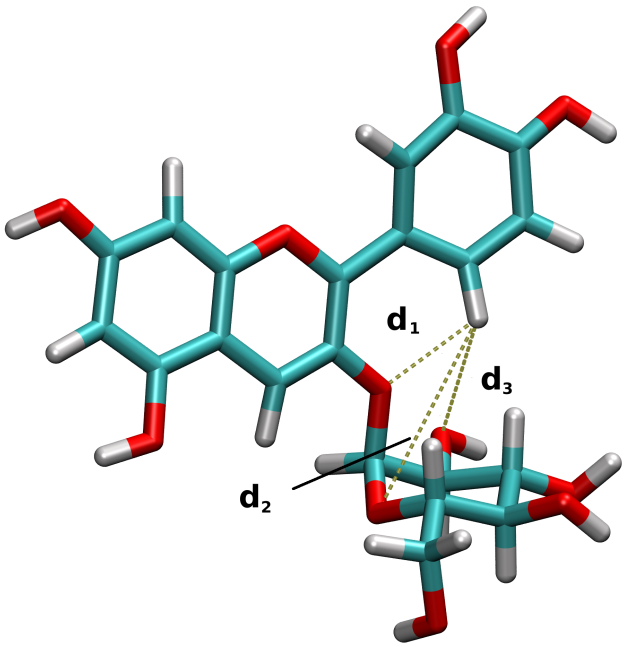


Figure 3.13: The distances which we examined to understand the asymmetries in the θ distribution among different protomers. (see the table 3.6 for detailed values of those bond lengths.

In order to ascertain the effects of this distortion onto the main absorption peak, we monitor the HOMO-LUMO (HL) gap along various AIMD trajectories. HL gaps are a convenient way to study the trends of optical transitions for a large number of configurations, because this gap is a lot easier to compute than the full absorption spectrum, while correlating very accurately with the position of the $S0 \rightarrow S1$ transition. HL gaps were calculated at PBE86 [62] level in presence of implicit solvent on 1000 geometries extracted from each AIMD trajectory, and were binned according to the θ values of the corresponding structure with a bin-width of 5°. The number of frames falling in each bin is of course θ-dependent and so is the statistical relevance. In this way, HL gaps can be tracked during the AIMD dynamics and computed from many configurations. We show here that for such chromophores with a single, intense, mostly H-L transition there is a

	d Å	C1	C2	C3	C4
$\mathbf{A}^{+}$	d_1	2.2 ± 0.4	2.3 ± 0.4	2.3 ± 0.4	2.3 ± 0.4
	d_2	4.2 ± 0.4	3.4 ± 0.6	3.2 ± 0.6	4.5 ± 0.5
	d_3	2.9 ± 0.6	4.4 ± 0.6	4.5 ± 0.5	3.4 ± 0.8
$\mathbf{A}^{\circ}_{4'}$	d_1	2.3 ± 0.4	2.1 ± 0.3	2.3 ± 0.3	2.3 ± 0.4
	d_2	3.2 ± 0.7	2.7 ± 0.5	2.7 ± 0.7	3.3 ± 0.7
	d_3	4.3 ± 0.8	4.2 ± 0.7	4.6 ± 0.4	4.3 ± 0.9
$\mathbf{A}^{\circ}_{7}$	d_1	2.2 ± 0.3	2.2 ± 0.3	2.2 ± 0.3	2.1 ± 0.3
	d_2	4.3 ± 0.3	3.0 ± 0.5	4.3 ± 0.5	3.4 ± 0.6
	d_3	2.9 ± 0.8	4.4 ± 0.6	3.3 ± 0.6	3.8 ± 0.8
$\mathbf{A}^{-}_{4'7}$	d_1	2.2 ± 0.3	2.2 ± 0.3	2.3 ± 0.3	2.1 ± 0.3
	d_2	3.8 ± 0.7	2.9 ± 0.6	2.8 ± 0.5	3.4 ± 0.5
	d_3	3.9 ± 0.6	4.5 ± 0.5	4.4 ± 0.5	3.7 ± 0.5
$\mathbf{A}^{-}_{4'5}$	d_1	2.4 ± 0.4	2.2 ± 0.3	2.3 ± 0.3	2.2 ± 0.4
	d_2	3.3 ± 0.7	2.5 ± 0.4	3.5 ± 0.7	3.8 ± 0.5
	d_3	4.8 ± 0.7	4.4 ± 0.6	4.0 ± 0.5	3.8 ± 0.7

Table 3.6: Average distances and standard deviations (Angstrom) between H1'/H6' and O1', O2' and O5' of the glycosyl moiety reported in d1, d2,d3, respectively for A^{+}, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$ and $A^{-}_{4'5}$. See figure **??**.

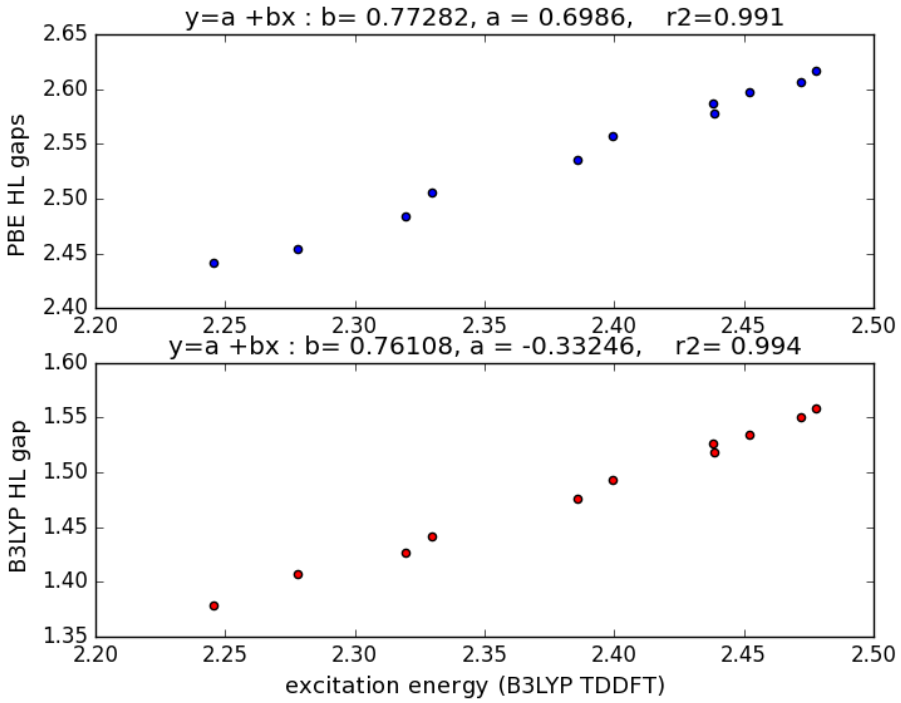


Figure 3.14: Correlation between HL gaps (calculated at both PBE (top y-axis) and B3LYP level (bottom, y-axis) and optical gap, i.e. excitation energy at TDDFT B3LYP level (x-axis) on structure from the relaxed scan at different value for *theta*. Linear regression parameters are also reported in the graph headers.

linear correlation of the HL gaps with the optical gap (vertical excitation energy $S_0 \rightarrow S_1$). Results are reported in Fig. 3.14.

In Fig. 3.15 we show the distribution of the θ dihedral for the five molecular species considered in this work (left panel), along with the average HL gap recorded within each θ bin for each conformer (right panel). The mid panel reports the average spectra recorded for each conformer. We notice that in most molecular states, a greater deviation from planarity results in a decreased HL gap. This trend however is not of the same magnitude across different species, and is indeed reversed in the case of A_7°. The nature of the frontier orbitals across the different molecular states is crucial to understand such trends. In order to get further insight, we utilize an indicator, B_π, of the bonding character along the C2–C1′ bond in the

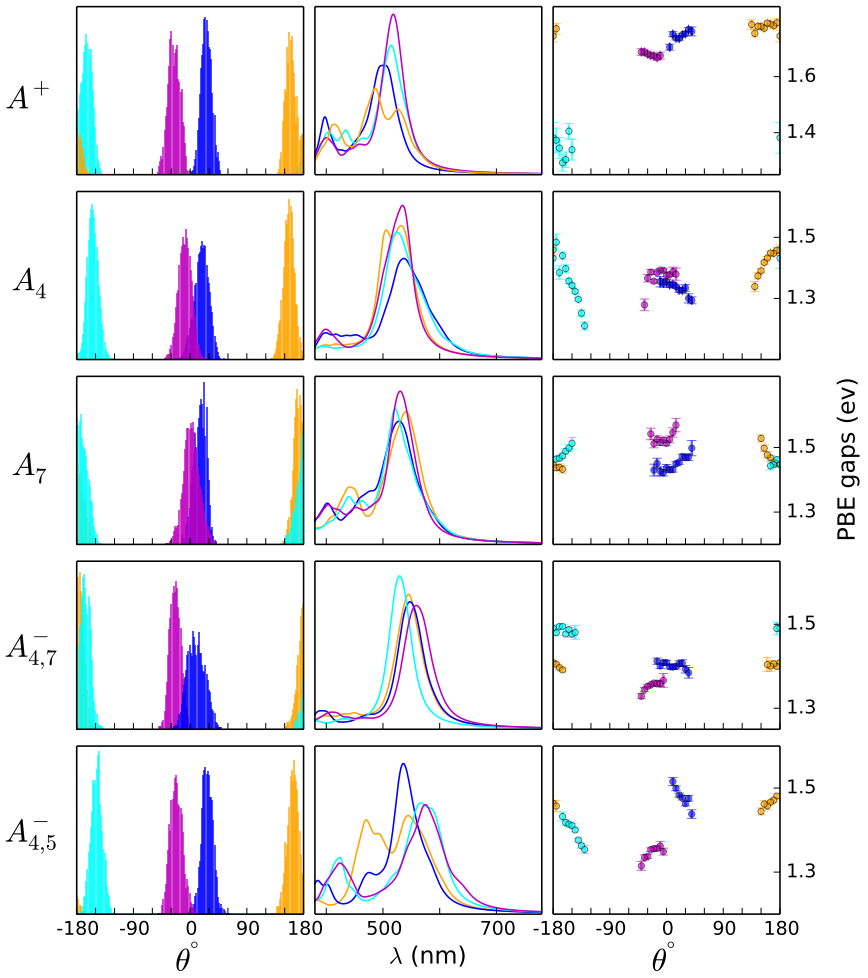


Figure 3.15: Results for A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$, and $A^-_{4'5}$ species of C3G are reported from top to bottom. Columns from left to right: histograms showing the distribution of the θ dihedral angle for all clusters equilibrated in ab initio MD (AIMD), spectra of each of the conformers, and HOMO LUMO gaps calculated for frames extracted from AIMD trajectories. All data referring to the C1, C2, C3, C4 conformers are reported in cyan, blue, pink and orange, respectively.

various frontier orbitals, defined as:

$$B_\pi(i) = \frac{|c_{i2} + c_{i1'}|}{\sqrt{2(c_{i2}^2 + c_{i1'}^2)}} \tag{3.1}$$

where c_{ix} is the projection of the i-th molecular orbital ($i = \{\text{HOMO}, \text{LUMO}, \cdots\}$) over the $2p_z$ atomic orbital of the x-th C atom ($x = \{1', 2\}$). The relative sign of c_{i2} *vs.* $c_{i1'}$ determines the resulting B_π: $B_\pi \sim 0$ reflects strong antibonding character, while $B_\pi \sim 1$ corresponds to a fully bonding π orbital. The computed values of B_π are shown in Table 3.7. This simple quantity is able to explain the behaviour of orbital energies upon distortion around C2 C1' bond: we expect the energy of a strongly bonding orbital to increase distortion, whereas an antibonding orbital should be stabilized by distortion. The resulting effect on the gap depends on the relative variation of the HOMO and LUMO energies. B_π(HOMO) follows the order (from least bonding to most bonding): $\mathsf{A}_7^\circ < \mathsf{A}^+ < \mathsf{A}_{4'7}^- \sim \mathsf{A}_{4'5}^- < \mathsf{A}_{4'}^\circ$; while for the LUMO the trend is: $\mathsf{A}_{4'}^\circ < \mathsf{A}_{4'7}^- \sim \mathsf{A}_{4'5}^- < \mathsf{A}^+ < \mathsf{A}_7^\circ$. This trend is reflected in the HL gaps *vs.* θ behavior reported in Fig. 3.15. For $\mathsf{A}_{4'}^\circ$ the optical gap markedly decreases with an increasing θ distortion, inducing a red shift in the absorption spectrum; for $\mathsf{A}_{4'7}^-$ this trend is slightly damped; in the case of A^+ the behavior is inverted and displays a slightly increasing trend with an increasing θ distortion; finally, in the case of A_7° the same trend displayed in A^+species is amplified resulting a blue shift in the absorption spectrum. Finally, for $\mathsf{A}_{4'5}^-$, based on B_π trends should be similar to $\mathsf{A}_{4'7}^-$ but we actually observe a more marked behaviour of HL gaps. This partial mismatch can be explained by the proximity of the HOMO-1 and HOMO energy levels in this species (see above), which is connected with the appearance of a second peak in the visible spectrum and partially undermines the single electron-hole pair picture of the optical transitions. We see that the bonding character depends on the protonation pattern more than the charge states—in fact both neutral protomers $\mathsf{A}_{4'}^\circ$ and A_7° are at the two extreme in bonding character for both HOMO and LUMO, and indeed the gaps clearly show opposite trends. The molecular orbitals reported in Fig. 3.12 qualitatively match B_π data.

The relative stability of different conformers characterized by different values of the θ dihedral, and the distribution of the θ within each conformer, are due to competing effects, namely electronic conjugation, which favors a planar geometry of the chromophore, and glycosylation at position 3, which exerts a torsion due to steric interactions between the sugar and the ortho ($2'$ or $6'$) hydrogen on the phenyl B ring [58]. Each molecular state has a

Frontier MO bonding characters $B_\pi(i)$ for C2–C1′					
	A^+	$A^\circ_{4'}$	A°_7	$A^-_{4'7}$	$A^-_{4'5}$
HOMO	0.44	0.98	0.08	0.77	0.78
LUMO	0.67	0.37	0.65	0.45	0.45
Bond order	1.17	1.29	1.15	1.24	1.23

Table 3.7: π-bonding character parameter $B_\pi(i)$, with respect to the C-C bond connecting the single and double rings for A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$ and $A^-_{4'5}$ of the frontier orbitals of C3G, as defined in equation 3.1. The the C-C bond order, calculated from the population analysis of the natural bond orbitals, is also shown. As expected, B_π(HOMO) mildly correlates with bond orders - both describing occupied levels.

different degree of conjugation, as shown by the variation in bond orders in Table 3.9, so the competition results in a different average value of θ.

3.3.3 Direct and indirect effects of explicit solvent on absorption spectra

As we have detailed in the previous paragraph the peak of the absorption spectrum highly depends on the θ distortion, but the extent of the fluctuations along θ are tightly determined by the relative orientation of the glucosyl substituent and the B ring. In this respect, an additional key element regulating the extent of cyanin distortion is the presence of explicit water molecules during the AIMD. In some conformers water molecules directly affect the relative orientation between the AC and B rings, by establishing a H-bonding water network that connect the O at the sugar C2 with the H(O) at 4′ of the chatecol B ring, or the H(O) at 5 (A ring) connected to the sugar with one H-bonded water. Moreover, persistent H-bonds are formed with the oxygens in position 3′, 4′, 5 and 7 of the C3G that strongly affect the peak of the absorption spectrum. Beside its explicit influence in selected cases, water molecules modulate the sugar position and in turn indirectly affect the optical properties.

As a control simulation to assess the role of water in tuning/stabilizing the most relevant conformers we performed AIMD simulations in vacuo

for all of them. Strikingly, during these simulations the position of α_6 assumes a wider ranges of values, which results in a larger exploration of the conformational space even along θ (see figure 3.16). This reveals explicit water molecules either by establishing an order wires of H-bond connecting different part of the molecule or by creating an ordered solvation shell around the hydrophilic moiety of C3G markedly contribute to stabilize the life-time of each conformer. This remarks the utmost importance of including explicit waters during static or dynamics QM studies in order to predict realistic structures for the spectrum calculations.

Furthermore, both the average and the spread in the distribution of θ is crucially affected by the solvent. The C2 and C3 conformers are characterized by the presence of an H-bonding wire with two water molecules connecting the hydroxyl at position $3'$ with the sugar hydroxyls. The presence of these rather persistent wires locks in the sugar configuration, thus tuning and locking the θ angle. This effect has been confirmed by comparing our explicit-solvent AIMD simulations with similar ones performed in vacuo, as shown in Fig. 3.16, where results for the C2 and C3 conformers are reported in pink and blue, respectively, as in Fig. 3.15. which we have confirmed by AIMD simulations of the conformers in vacuo. As shown in figure In Fig. 3.16. We notice that water has a restraining effect on α_6, whose values fluctuate more freely in vacuo, thus resulting in a larger exploration of the conformational space also along θ. Remarkably, the most likely values of α_6 values differ in the water and in vacuo. This reveals that the water molecules shape the conformational landscape of each cyanin species by creating an ordered solvation shell around the hydrophilic moiety of C3G.

We have also performed QMMM simulations using QMMMW (a wrapper for QM/MM simulations with QUANTUM ESPRESSO and LAMMPS [69]), keeping C3G as a QM part and including waters as a mm part to check if the presence of MM water is enough to obtain the reasonable geometries for the solute. Comparing, dihedral angle distributions of C3G trajectories coming from QMMM and CP do not match each other. Nevertheless, the dihedral distributions are not as broad as in case of vacuum meaning that classical waters restrain molecule to explore a wider range of angles as during CP dynamics. However the peaks of the distributions do not match the ones coming from ab initio molecular dynamics trajectory. Thus, we can assume that QMMM simulations are not accurate enough to reproduce the correct configurations of the C3G for spectrum calculations.

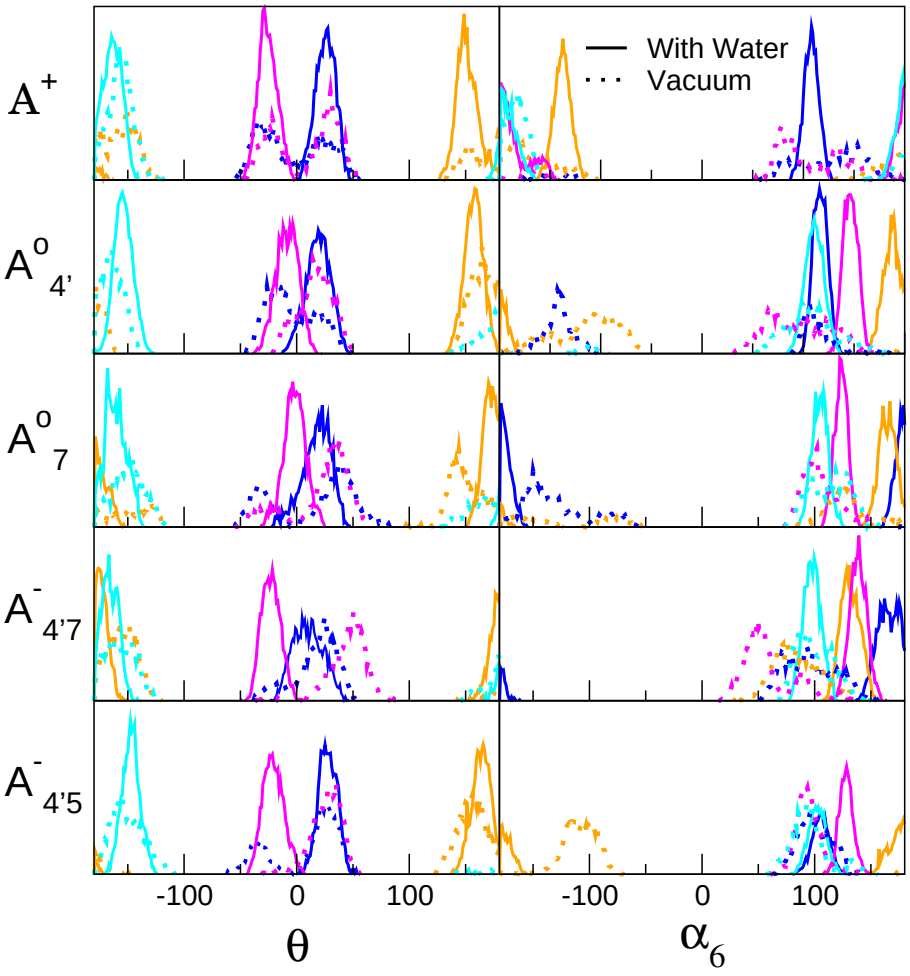


Figure 3.16: Comparison between the θ and α_6 dihedral-angle distributions resulting from AIMD simulations for C3G in explicit solvent (continuous lines) and in vacuo (dotted lines). The color code refers to different conformers as in Fig. 3.15. In particular, results for the C_2 and C_3 conformers are reported in blue and pink, respectively.

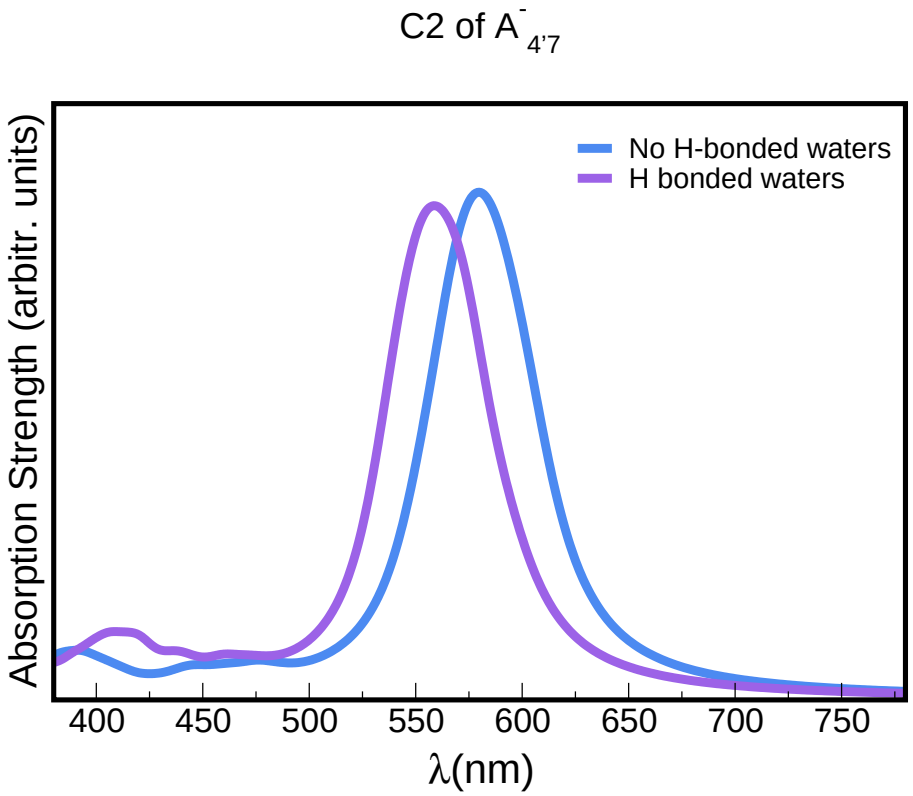


Figure 3.17: Comparison between the two averaged spectrum, both done by averaging TDDFT spectum on the same solute frames chosen from AIMD trajectories, one including H-bonded waters and the other without solvent molecules. We can see the difference of two spectra on the figure. Moreover, color code corresponds to the colors they express. As shown in the figure the difference is very obvious, making it neccessary to include H-bonded waters in spectrum calculations.

In order to understand the direct effect of persistent H-bonded water on absorption spectra we have computed the average spectrum with and without those explicit water molecules for various comformers. The difference in absorption spectrum peak λ_{max} between those two cases is significant(8-10 nm for A^+, 4-7 nm for $A^\circ_{4'}$ and A°_7 neutral species and is the most significant for negative $A^-_{4'7}$ and $A^-_{4'5}$ species (12-20 nm). In figure 3.17 the most extreme case is presented, C2 conformer of $A^-_{4'7}$, where the peaks of two averaged spectrum differ by 20 nm and the colors on the figure represent the real colors to which spectra correspond. We can clearly see the importance of including the H-bonded waters during the TDDFT calculations.

3.4 Generalization to other anthocyanins

In an attempt at generalizing the relationship between internal chromophore distortions and optical properties we have also considered the anthocyanins pelargonin and delphinin, which differ from C3G in the number of hydroxyl groups on the B ring (see chapter 1). Relaxed scans of the θ angle (by optimizing all other degrees of freedom of the system and fixing θ at different values) for all investigated charge states and tautomer of C3Gs as well as for pelargonidin and delphinidin were performed with QUANTUM ESPRESSO at PBE86 [62] level in presence of implicit solvent [64]. In these scan the glucosyl moiety was replaced by a methyl group to remove the steric effect of the sugar, but to still account for its electronic properties. Geometry optimizations of those reduced model systems, with θ constrained at different values, (relaxed scan), reveal that, independently of the charge state and of the protomeric form (Fig. 3.18), the calculated HL gaps show that delphinidin assumes the bluest nuance, followed by cyanidin and pelargonidin, in line with experimental findings [70, 71] and theoretical investigations [10]. Consistently with cyanin, the sensitivity of the HL gap to the θ distortion is confirmed. Indeed, for the same arguments of C3G, a decrease of the gap along with an increased internal distortion is again more pronounced in the neutral quinonoid form deprotonated at $4'$ position, and damped when deprotonation occurs at positions 7 or 5 (Fig. 3.18).

As a result, we add an additional element to this entangled mixture of electronic/structural traits: the polarization induced by the number of hydroxyl substituents on the B ring also contribute to tune the chromophore optical properties. This takes place by virtue of the electron withdrawing effect of the hydroxyl substituents, which in turn affects the relative energy of the HL gap. Cyanin, delphinin and pelargonin are the most abundant species of the anthocyanin family (90%), thus understanding the behaviour of these molecules we can assume that our results can be generalized to the all family of anthocyanins.

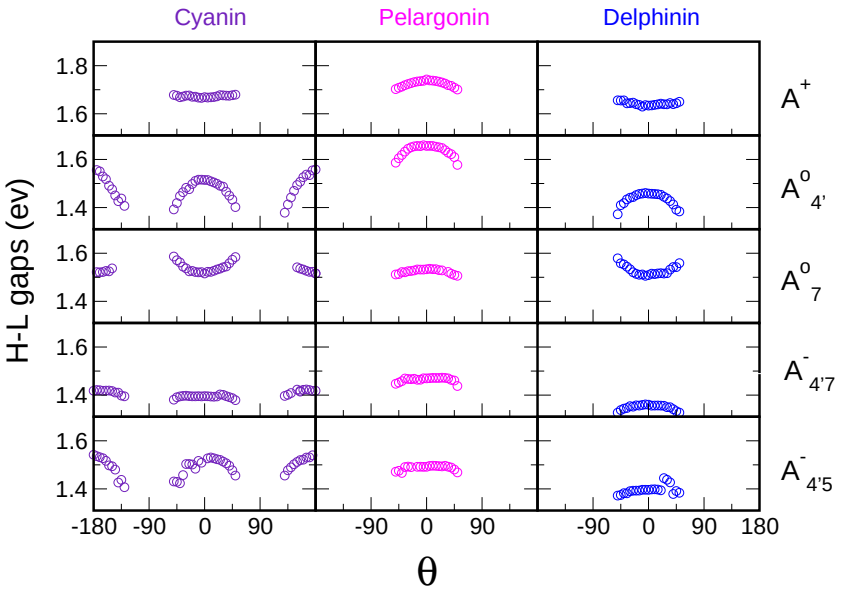


Figure 3.18: The trend of the HL gaps with respect to the θ dihedral angle, for a reduced model system of antocyanine. θ results to be a key internal degree of freedom for all charge states of investigated molecules.

Bond Order					
Conformer Name	$\mathbf{A^+}$	$\mathbf{A^\circ_{4'}}$	$\mathbf{A^\circ_7}$	$\mathbf{A^-_{4'7}}$	$\mathbf{A^-_{4'5}}$
C1	**1.19**	**1.28**	**1.13**	**1.25**	**1.21**
C2	**1.16**	**1.29**	**1.16**	**1.22**	**1.22**
C3	**1.16**	**1.28**	**1.15**	**1.24**	**1.23**
C4	**1.15**	**1.29**	**1.15**	**1.24**	**1.24**

Table 3.8: Bond order parameters of the C-C bond connecting the single and double rings for all most representative conformers of A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$ and $A^-_{4'5}$.

Bond Order					
	$\mathbf{A^+}$	$\mathbf{A_4}$	$\mathbf{A_7}$	$\mathbf{A^-_{4,7}}$	$\mathbf{A^-_{4,5}}$
Cyanin	**1.17**	**1.29**	**1.15**	**1.24**	**1.23**
Pelargonin	**1.18**	**1.30**	**1.16**	**1.25**	**1.25**
Delphinin	**1.16**	**1.27**	**1.15**	**1.23**	**1.23**

Table 3.9: Bond order parameters of the C-C bond connecting the single and double rings for A^+, $\mathsf{A}^\circ_{4'}$, A°_7, $\mathsf{A}^-_{4'5}$and $\mathsf{A}^-_{4'7}$of Cyanin, Pelargonin and Delphinin. *Bond order* calculations were performed using the Natural Bond Order (NBO) frame on single point `Gaussian09` calculations on selected snapshots along the AIMD dynamics (same used for spectrum).

3.5 Conclusion

In this work we have considered the different anthocyanin species formed at various pH ranges in water solution. The sophisticated computational protocol employed here enabled us to consider both thermal vibrations and the explicit solvent medium, accounting for the slow conformational properties of ANTs at finite temperature.

We stunningly reveal that the relative twist, and to a minor extent the bend, of two dihedral angles in the molecule are key ingredients for the color development of this anthocyanin in different forms. Surprisingly, the extent of single/double character of the bond connecting the favylium and catechol rings tightly regulate the degree of twisting and bending of C3G, which is in turn responsible for the excitation energy of the molecule and thus for the resulting absorption spectrum.

This study highlights that the conformational properties of molecules at finite temperature, as well as thermal vibrations and the effect of explicit solvent molecules are critical to dissect the molecular traits responsible for tuning the optical properties of solvated dyes.

This intricate connection between structural/electronic and optical prop-

erties is also shared by other members of the anthocyanin family such as pelargonin and delphinidin.

Chapter 4

Metal Complexation

In this chapter, I introduce the study that further extends our work on cyanidin-3-glucoside discussed in chapter 3 by assessing the impact of metal complexation on its structural, dynamical, electronic, and optical properties. To this aim we study:

- *The conformational effects of* Al^{3+} *complexation on C3G.*
- *The effect of* Al^{3+} *on different charge states of C3G.*
- *The effect of different ratios of* Al^{3+}*:C3G (1:1; 1:2 and 1:3).*

As it has been mentioned in chapter 1, pure anthocyanins are characterized by poor stability at weakly acidic conditions. Experimental studies show [72–77] that complexation of anthocyanin contributes to increase their stability in solution [4]. With metal complexation, metal ions (Me^{n+}) bind to the chromophore of anthocyanins. As a result, the complex is protected from nucleophilic attack, which in turn prevents degradation of anthocyanin with increasing pH. Furthermore, metal chelation by anthocyanin is thought to increase the color expression of the pigments [78]. Thus, complexation of anthocyanins with metal ions seems to have the potential to alleviate some of the adverce effects that limit the large-scale deploiment in the food industry.

Formation of metal-anthocyanin complexes are widely observed in Nature. However, in order to chelate metal ions and undergo a color change in solution, the chromophore must bear at least two hydroxyl groups on

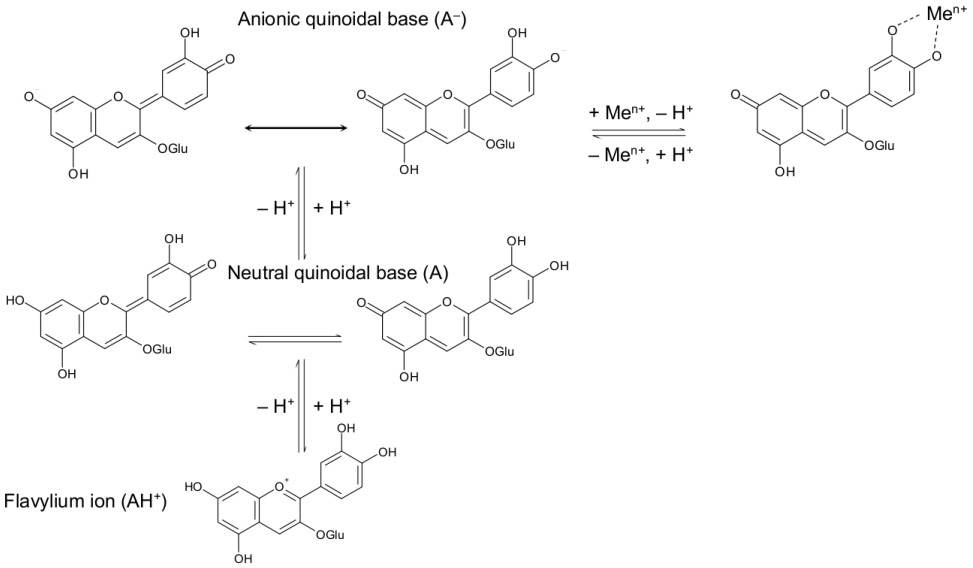


Figure 4.1: Structures of C3G at different pH values and their possible complexation with metals.

the B ring [77]. Not all the anthocyanins are able to form chelates with metal ions, so the most common compounds (see figure 1.4) mentioned in chapter 1, should be divided into two groups. The first group comprises anthocyanins with two or three hydroxyl substituants on the B ring, which are cyanidin, delphinidin and petunidin, (as well as their derivatives after glycosylation). These compounds usually bind to metal ions by two OH substitutes in the ortho-position in the B ring. Conversely, anthocyanins with a single hydroxyl substituent (Pelargonidin), a single hydroxyl substitute and one methoxyl side group (Peonidin), a single hydroxyl substitute or two methoxyl side groups (Malvidin) on the B ring belong to the second group (as well as their derivatives after glycosilation) (see figure 1.4). Members of the second group do not form complexes with metal ions with a single anthocyanin.

A crucially important feature of anthocyanins, which distinguishes them from other compounds and favors their binding with metal ions consists in the pH-dependent dynamic equilibrium of their different structural forms in water solutions. As we have discussed in the previous chapter, for C3G at highly acidic conditions (pH $<$ 3) the flavylium cation predominates, expressing red color in solution, whereas at slightly acidic conditions (pH 3-5) series of chemical transformations occur. The colored species at that

pH ranges are the neutral quinoidal bases, formed after proton dissociation from flavylium cation. Those species give rise to pinkish colors (see chapter 3). At neutral/basic pH (6-9) anionic quinoidal bases are predominant, which express purplish colors in solution. metal complexation shifts the equilibrium amongst different structures at distinct pH values. Metal complexation favors the formation of purple-bluish coloring quinoidal form, and stabilize the resulting metal-anthocyanins complexes. As a consequence, at a given pH value both the position and the intensity of the basic absorption maximum changes. The final color expressed by the pigment is shifted from pinkish-purple to purplish-blue. The molecular mechanisms underlying this color shift remain unclear, despite being of key interest for many industrial companies searching different hues of blue color among the natural colorants.

According to the previous experimental studies, multivalent metal ions such as Al^{3+} and Fe^{3+} induce bathochromic shifts on the visible absorption of anthocyanins, developing blue colors when present in solution [79]. Unlike blue metalloanthocyanins, widely spread in blue flowers [4] that have fixed stoichiometric ratios of anthocyanin to flavones to metal ions (6:6:2), metal-anthocyanin chelated complexes seem to exist in equilibrium with varying degrees of anthocyanins association from single molecules to complexes of 3 molecules [79].

Until now, no systematic study has been done on the exact metal-ligand ratio at each pH interval, and even the information concerning the species existing at a different pH conditions is unclear. Based on the accuracy of our protocol, as demonstrated in previous chapters this study aims at understanding the species existing at various pH ranges, by comparing our theoretical results with experimental data.

4.1 Al^{3+}:C3G ; Study of 1:1 complexes

For the time being we focus on Al^{3+}:C3G complexation. Aluminum(III) is an octahedrally hexacoordinated ion that depending on pH can bind one, two, or three C3G molecules forming 1:1, 1:2, and 1:3 complexes, respectively. As a ligand, we have chosen C3G, since we have already studied it (chapter 3) and we can compare results obtained with and without metal complexation to better understand the direct effect of metal

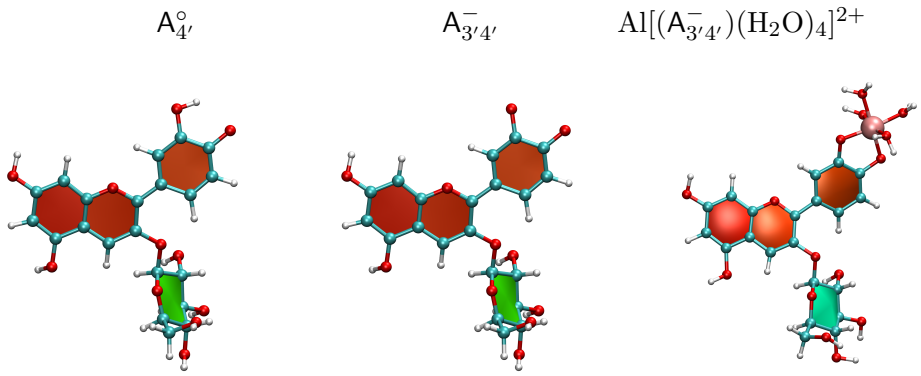


Figure 4.2: Left: $A^{\circ}_{4'}$, where position 4' is already deprotonated (exists at pH 3-6); middle $A^{-}_{3'4'}$ is formed after removal of H^+ from position 3' (this species does not exist in solution alone, but only before metal binding. Thus this is an artificial intermediate structure between pure anthocyanins and metal anthocyanin complexes.); right: 1:1 $A^{-}_{3'4'}$:Al^{3+} complex, Al^{3+} binding with oxygens at positions 3' and 4'. 4 other water molecules complete in the coordination sphere of Al^{3+}.

ions on the absorption spectrum. C3G has two hydroxyl substituant on the B ring allowing metal complexation to occur and is, in terms of frequency and abundance, the most common anthocyanin in nature.

Both theoretical [73,80] and experimental studies [72,81] have indicated that the 3′-4′ site of the B ring is the preferred Al^{3+} binding site (see figure 1.2). Metal binding is possible only when both hydrogens at 3′ and 4′ positions are removed from the chromophore. Indeed, it is well known [73] that most compounds bearing phenolic hydroxyl groups undergo metal complexation reactions upon deprotonation and metal binding at the same oxygen atom [73]. The main driving force for binding metal ions in such complexes is the large binding energy in the complex compared with the binding energy of the H^+ ion [73]. We suppose that the same reason stands also for complex formation between C3G and Al^{3+}.

The formation of the stable complex at slightly acidic conditions between Al^{3+} and C3G is favored by Al^{3+} facilitating the loss of the hydrogens from the hydroxyl groups at positions 4′ and 3′ (see figure 1.4), transforming the flavylium cation (A^+) first to neutral species and then to a anionic quinonoid base [76, 79]. Since Al^{3+} in water solution is normally coordinated to 6 water molecules, anthocyanin replaces two of these molecules, still resulting in an octahedral coordination. (See figure 4.2).

In order to explore the optical properties of C3G in the presence of Al^{3+}

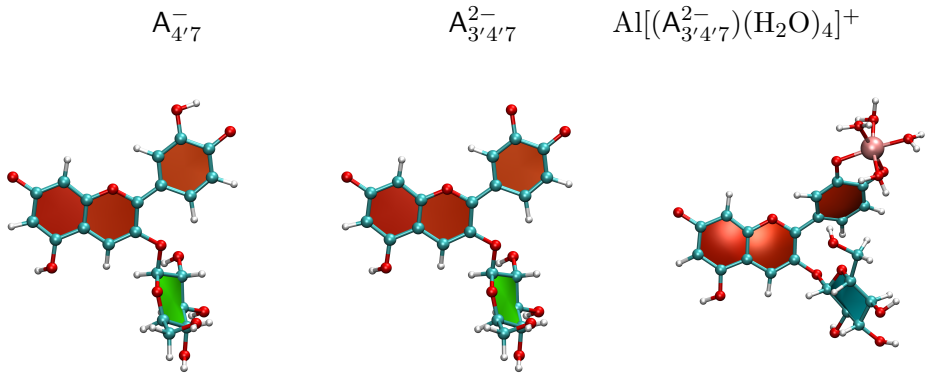


Figure 4.3: Left: $A^{-}_{4'7}$, (exists at pH 6-9 ranges); middle $A^{2-}_{3'4'7}$ is formed from $A^{-}_{4'7}$ after removal of H^{+} from position 3' (like in case of $A^{-}_{3'4'}$ also $A^{2-}_{3'4'7}$ does not exist in solution); right: 1:1 $A^{-}_{4'7}$:Al^{3+} complex, Al^{3+} binding with oxygens at positions 3' and 4'. 4 other water molecules are in the coordination sphere of Al^{3+}.

ions and to evaluate their spectrum and color accurately, explicit-solvent BO MD simulations were performed for each one of the four most populated conformers (C1, C2, C3, C4) of $A^{\circ}_{4'}$ and $A^{-}_{4'7}$ in a complex with Al^{3+}. For ab initio dynamics a mixed Gaussian and plane waves approach supported by the CP2K [82] software package was used, with DZVP-MOLOPT-SR-GTH basis set and GTH-BLYP pseudopotential from CP2K public repository. The number of explicit water molecules was ~ 300, so as to fit a simulation cell of linear dimensions ~ 30 Å. An MD time step of $\sim$0.5fs was used. The simulation box including explicit solvent was pre-equilibrated via classical MD at constant pressure/temperature, while keeping the Al:anthocyanin complex restrained. AIMD simulations were finally performed at constant volume and (room) temperature (NVT) using the CSVR thermostat [83–86] for at least 15 ps per equilibrated trajectory.

The $A^{\circ}_{4'}$ and $A^{-}_{4'7}$ species forming complexes with metal ions have to loose one extra H^{+} from position 3′, so that in the metal anthocyanin complexes with $A^{\circ}_{4'}$ and $A^{-}_{4'7}$, the ligand carries a charge of -1(C3G deprotonated at positions 3′ and 4′ - $A^{-}_{3'4'}$) and -2 (C3G deprotonated at positions 3′, 4′ and 7 - $A^{2-}_{3'4'7}$), respectively. Thus, Al bearing a +3 charge, the resulting complexes exhibit a charge of +2 and +1, respectively. Hereafter these complexes will be indicated as $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$ and $Al[(A^{2-}_{3'4'7})(H_2O)_4]^{+}$, respectively.

Optical spectrum were computed using the localized basis set package

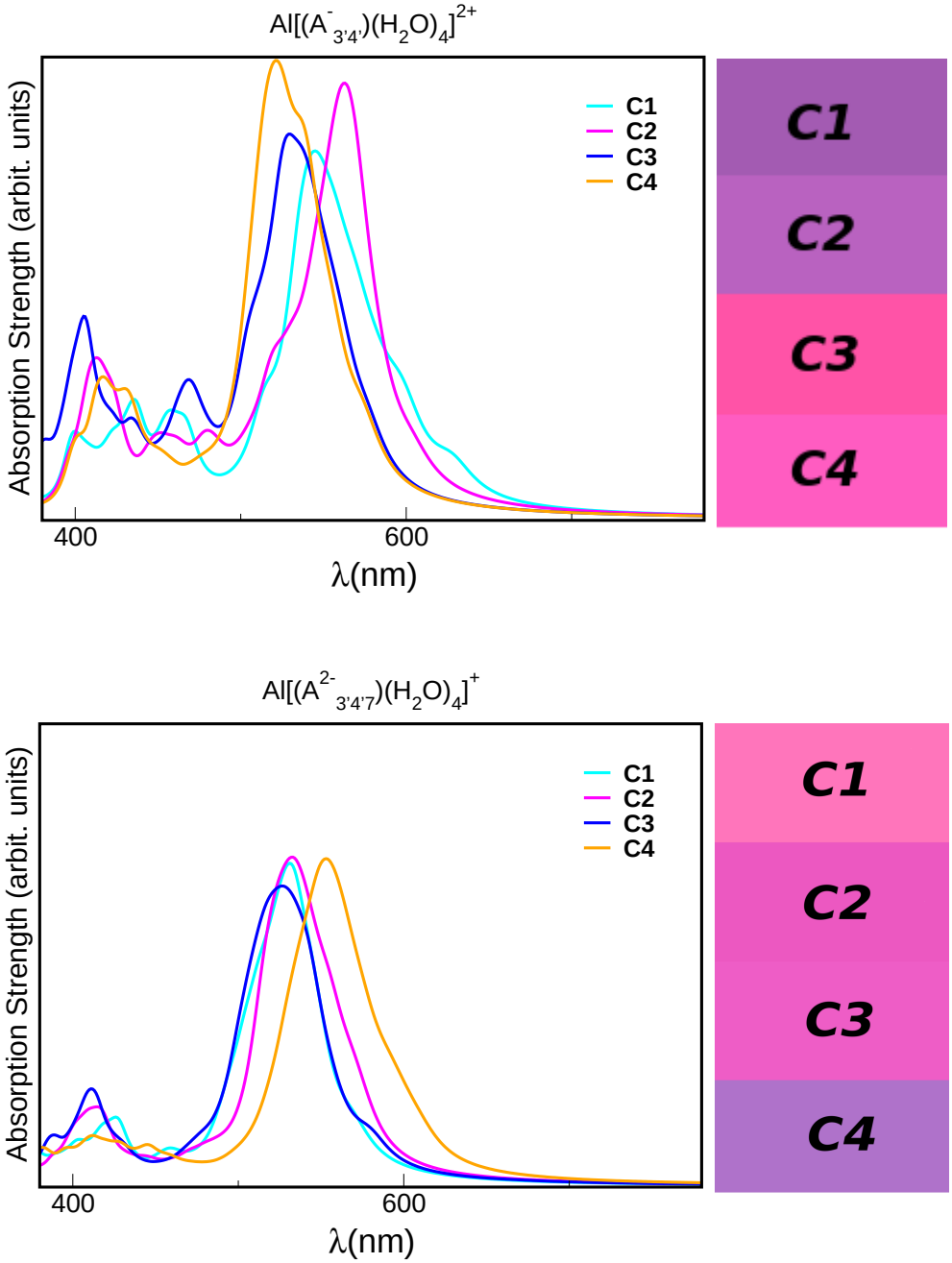


Figure 4.4: Top Panel: Left: the average absorption spectrum along the AIMD trajectory for conformers C1, C2, C3, C4 of $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$, Right: the colors corresponding to each conformer in water solution. Bottom Panel: The same but for $Al[(A^{2-}_{3'4'7})(H_2O)_4]^{+}$ species.

gaussian [53] with B3LYP functional [48] and basis set 6-311g(d,p). Molecular configurations for TDDFT calculations were extracted from the AIMD trajectory at every 0.5 ps, ensuring statistical independence. During the calculation of the spectrum only a few water molecules, which were found to be persistently H-bonded to the chromophore, were explicitly considered. The rest of the solvent molecules were accounted for through an implicit solvent (PCM). Finally, the spectrum of a given AIMD trajectory was computed as the average of the spectra of each selected frame.

This method of computing the average spectrum was applied to C1, C2, C3 and C4 conformers of $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$ (figure 4.2) and $Al[(A^{2-}_{3'4'7})(H_2O)_4]^{+}$ (see figure 4.3)complexes. The final average spectrum for each of the mentioned cases and their expressed colors in solution is shown in figure 4.4. For both $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$ and $Al[(A^{2-}_{3'4'7})(H_2O)_4]^{+}$, the most intense transition of the absorption spectrum is between the HOMO and the LUMO, like it was the case for C3G. Nevertheless, the absorption peak λ_{max} is redshifted causing an overall bluing of the solution. It is worth to notice that the difference between the spectra of each configuration is significant, confirming that the study of complete configurational space is very important for accurately predicting the color of a given molecule.

4.2 1:1 Al^{3+}:C3G complexes; effect of complexation on different charge states

In order to clarify the effects of aluminum complexation on C3G, we compare the spectrum of each conformer computed before and after complexation with Al^{3+}. The aim of this analyses is to explore the effect of Al^{3+} on the absorption spectrum, caused by the geometrical deformations occurring upon metal complexation of anthocyanins.

In figure 4.5 the spectrum with their corresponding colors of four conformers C1, C2, C3, C4 of $A^{\circ}_{4'}$ and $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$ are displayed. The effect of complexation on conformers C1 and C2 is significant, the spectrum is redshifted by 20-30 nm, resulting into a marked change of expressed color. Conversely, the optical properties of the C3 and C4 conformers are not strongly influenced by metal binding. There is a slight blueshift of λ_{max} in case of C3, resulting very small change in color. Similarly, in case of C4 the spectrum is slightly sharper, and no significant variation in color

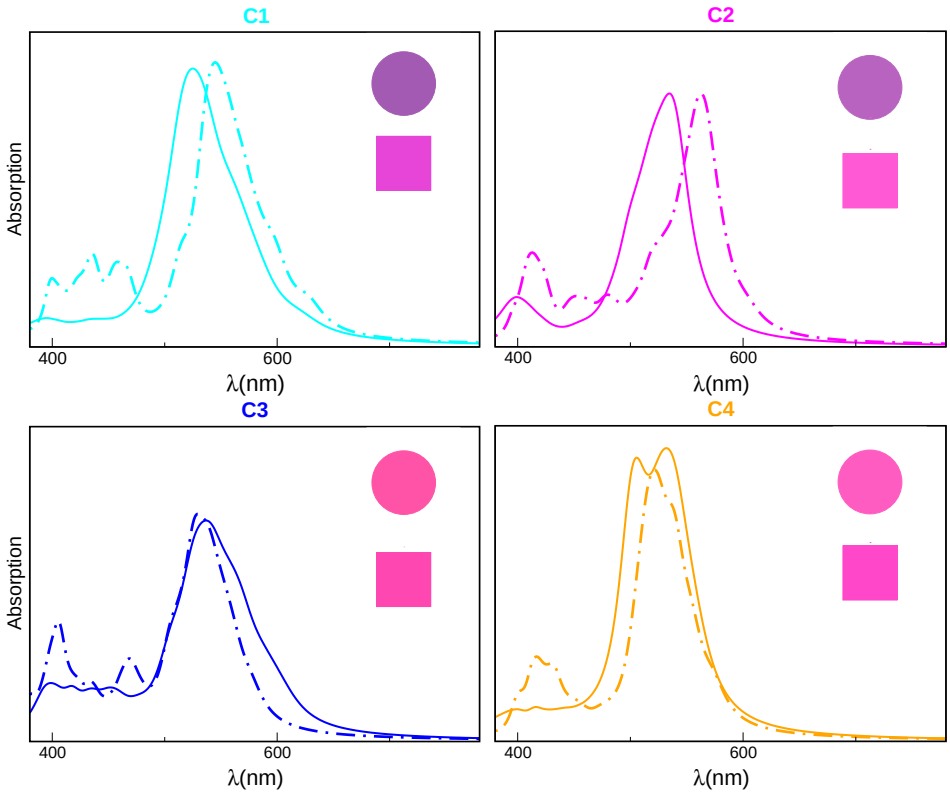


Figure 4.5: Four plots representing results for four different conformers of neutral A_4°, species of C3G C1, C2, C3, C4 and their complexation with Al^{3+}. The straight line is the average spectrum of a given conformer along AIMD trajectory, dashed line is the average spectrum of the same conformer in presence of Al^{3+} ion along AIMD trajectory. The color of the square represents the color of the pure C3G (of a given conformer) in water solution and the color of the circle represents the color of the same conformer of C3G with Al^{3+} complex in water solution.

expression is observed.

At first, we have tried to understand the effect of the Al^{3+} ion on the geometry of each conformer.In figure 4.6 we report the distribution of θ and α_6 dihedral angles along the AIMD of pure $\mathsf{A}^{\circ}_{4'}$ and $Al[(\mathsf{A}^{-}_{3'4'})(H_2O)_4]^{2+}$. From the α_6 dihedral distribution we can see that in the C3 conformer, the sugar position changes the most upon complexation, resulting in a drastic change of the θ distribution as well. As a result, θ is very distorted, when C3G is free in solution and becomes planar upon binding to the metal. This explains the blueshift in spectrum observed for the C3 conformer, agreeing with our previous results that conformers bearing the most distorted θ configurations display the most redshifted spectrum, (see chapter 3). In the case of conformers C1 and C2, in the presence of aluminum ion θ twistes more than for pure C3G, resulting in a redshift of the overall spectrum in both cases. Conversely, in case of conformer C4 no significant conformational changes occur and the resulting spectrum of $Al[(\mathsf{A}^{-}_{3'4'})(H_2O)_4]^{2+}$ complex is similar to that of the corresponding of $\mathsf{A}^{\circ}_{4'}$ conformer.

We have found that the main changes in the spectrum upon metal complexation with $\mathsf{A}^{-}_{3'4'}$ come from the geometry deformation caused by its binding to the Al^{3+} ion. Additionally, we think that the main cause for geometry changes is not the metal ion itself but the water olecules coordinated to it. These create a bulky group affecting the relative distribution of θ and the other dihedrals describing the orientation of the sugar.

We have carried out the same analysis for $\mathsf{A}^{-}_{4'7}$ and $Al[(\mathsf{A}^{2-}_{3'4'7})(H_2O)_4]^{+}$ species by comparing the spectrum of each conformer computed before and after complexation with Al^{3+}. The spectrum with their corresponding colors of four conformers C1, C2, C3, C4 of $\mathsf{A}^{-}_{4'7}$ and $Al[(\mathsf{A}^{2-}_{3'4'7})(H_2O)_4]^{+}$ are displayed in figure 4.7. The dihedral distributions for all conformers of those two species are shown on figure 4.8. The geometries of the conformers C3 and C4 undergo significant changes. Sugar orientation is completely changed for those two conformers, resulting in less θ distortion for conformer C3, and more distortions for conformer C4, hence, causing blueshift and redshift in average spectrum, respectively. Also, in the case of C2, the geometry is slightly affected by the metal complexation, however, the spectrum is very blueshifted. Thus, another degree of freedom must be influenced by the Al^{3+}, causing additional changes in the absorption spectrum. C1 is the only conformer whose spectrum does not show almost any shift.

As a final check, we wanted to make sure that the geometry changes

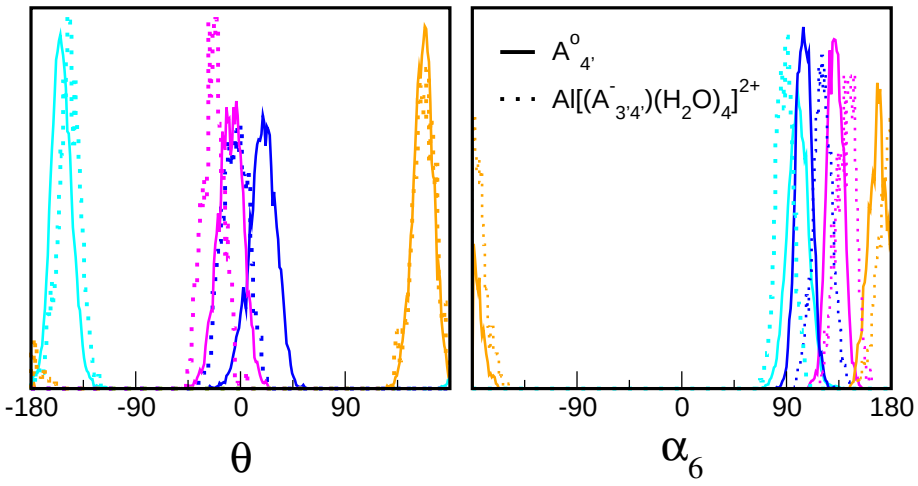


Figure 4.6: θ and α_6 dihedral angle distributions coming from AIMD dynamics. Straight lines correspond to the conformers C1, C2, C3, C4 of $A^{\circ}_{4'}$, dotted lines correspond to the same conformers of $A^{-}_{3'4'}$ in complex with Al^{3+}.

observed here were really the consequence of metal complexation and not the removal of H^+ proton at position 3′, which is usually retained by the C3G even at high pH values. It is reasonable to assume that proton removal from position 3′ occures only when the metal ion binds to C3G (either already deprotonated or initiating its deprotonation at position 4′) and by aproaching position 3′, it facilitates the second proton release. As soon as the hydrogen is removed, the metal ion binds to the *"lone"* oxygen. For this reason, we have run BOMD dynamics in the same conditions as discussed in the section 4.1 for the hypothetical species $A^{-}_{3'4'}$ and $A^{2-}_{3'4'7}$. In figure 4.9 we display the comparison of θ dihedral distributions coming from BOMD. On one side we compare $A^{\circ}_{4'}$, $A^{-}_{3'4'}$ and $A^{2-}_{3'4'7}$ and on the other side $A^{\circ}_{4'}$, $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$ and $Al[(A^{2-}_{3'4'7})(H_2O)_4]^{+}$ species. The distributions of $A^{\circ}_{4'}$ and two artificial species ($A^{-}_{3'4'}$ and $A^{2-}_{3'4'7}$) are almost identical, centered around θ 18° and having similar width. In contrast, the distributions in $A^{-}_{3'4'}$ and $A^{2-}_{3'4'7}$ upon complexation display a significant change in θ distribution compared to the initial structure $A^{\circ}_{4'}$. Hence, the conformational changes are induced by the presence of metal ions and, possibly, by water molecules in the first coordination shell, while deprotonation at position 3′ plays no major role in the geometrcal arrangement of the complex.

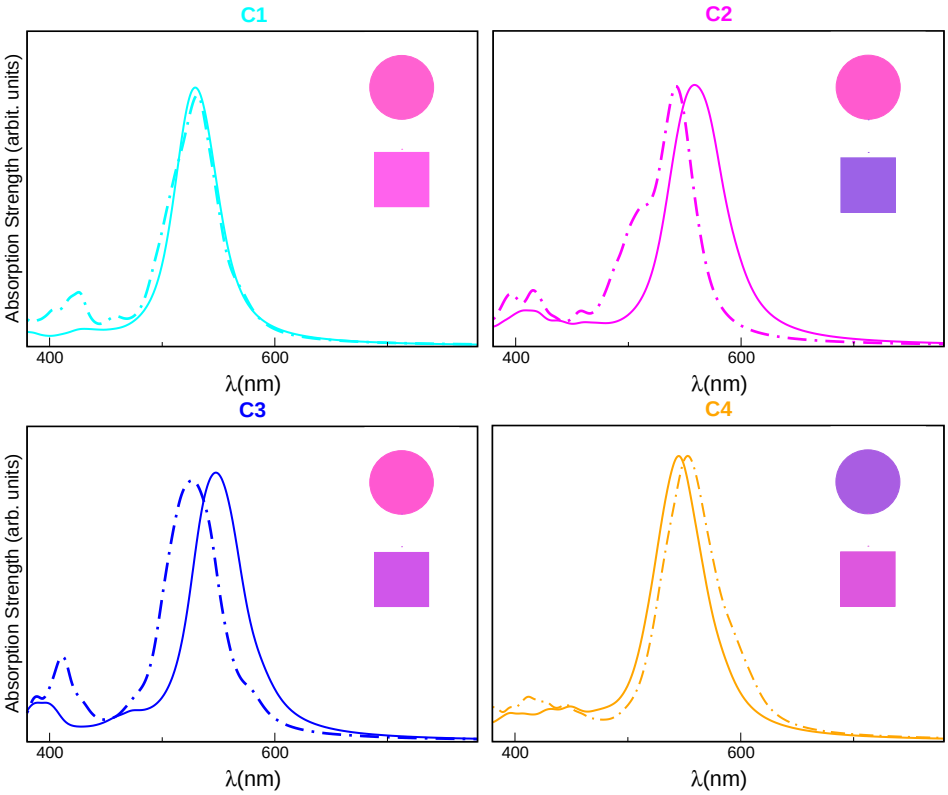


Figure 4.7: Four plots representing four different conformers of neutral $A^{-}_{4'7}$ species of C3G C1, C2, C3, C4, straight line is the average spectrum of a given conformer along AIMD trajectory, dashed line is the average spectrum of the same conformer (also deprotonated at position 3') with Al^{3+} ion bound along AIMD trajectory. The color of the square represents the color of the given conformer in water solution and the color of the circle represents the color of the given conformer of C3G in complex with Al^{3+} in water solution.

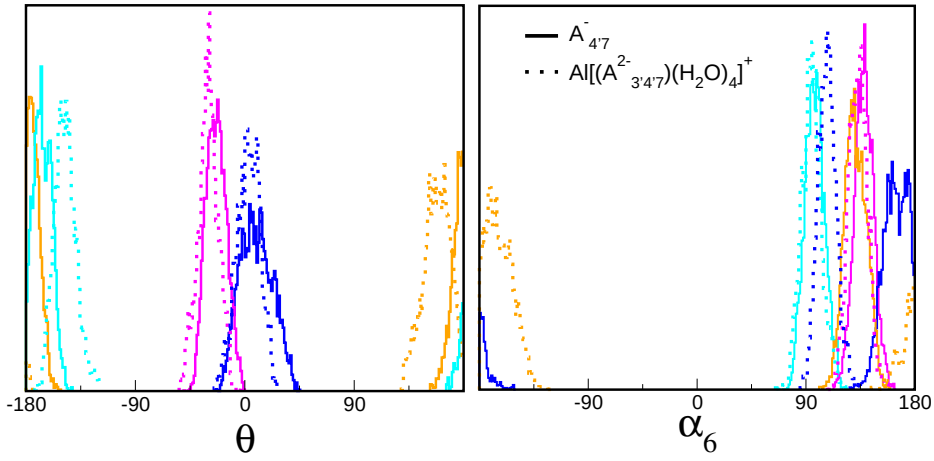


Figure 4.8: θ and α_6 dihedral angle distributions coming from AIMD dynamics. Straight lines correspond to the conformers C1, C2, C3, C4 of $A^{-}_{4'7}$, dotted lines correspond to the same conformers of $A^{2-}_{3'4'7}$ in complex with Al^{3+}.

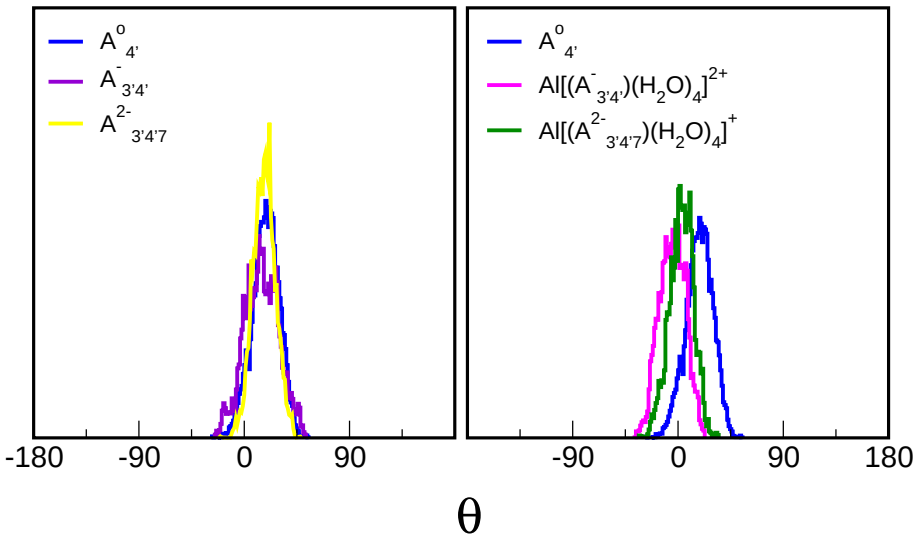


Figure 4.9: θ dihedral distributions coming from AIMD dynamics. Left: of $A^{\circ}_{4'}$, $A^{-}_{3'4'}$ and $A^{2-}_{3'4'7}$; Right: $A^{\circ}_{4'}$, $Al[(A^{-}_{3'4'})(H_2O)_4]^{2+}$ and $Al[(A^{2-}_{3'4'7})(H_2O)_4]^{+}$ species.

4.3 1:2 and 1:3 complexes

Experimental and theoretical [72, 75, 79, 80] work suggests that metal ions can bind to two or even three anthocyanin molecules at neutral conditions (pH 5-9) [75]. However, these studies do not elucidate the overall charge of the resulting complexes (the protonation state of anthocyanins).

Because of the computational cost involved, we have studied only 4 of these complexes. Two of them forming 2:1 complexes ($Al[(A^{-}_{3'4'})_2(H_2O)_2]^{+}$ and $Al[(A^{2-}_{3'4'7})_2(H_2O)_2]^{-}$ see figure 4.10) and two complexes in which the metal binds to three anthocyanins ($Al[(A^{-}_{3'4'})_3]$ and $Al[(A^{2-}_{3'4'7})_3]^{3-}$ see figure 4.11). For each of those species we have performed BOMD dynamics in explicit water solution (800-1000 explicit water molecules; For more details about the dynamics see section 4.1) and the absorption spectrum was computed as described in section 4.1. The average spectrum for each species and their expressed color are shown in figures 4.12, 4.13, 4.14, 4.15. From these figures we can clearly see that each of the spectra has two very broad peaks in contrast to the case of pure C3G or complexes in which the metal anthocyanin ratio was 1:1. The first peak is around 570-630 nm in case of 1:2 complexes and around 640-680nm for 1:3 cases, contributing to the blue color of the complex. The second peak is instead due to the interactions between the different chromophores in the complex. This lies around 450 nm and is the most intense for $Al[(A^{-}_{3'4'})_2(H_2O)_2]^{+}$ and $Al[(A^{-}_{3'4'})_3]$ cases, resulting greenish colors for those species. For $Al[(A^{2-}_{3'4'7})_2(H_2O)_2]^{-}$ and $Al[(A^{2-}_{3'4'7})_3]^{3-}$ complexes the second peak is around 490 nm. However, this peak being less intense than the first one, the overall color for those species are bluer.

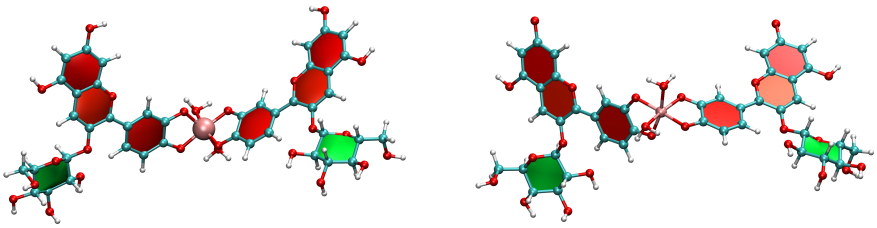


Figure 4.10: Left: $Al[(A^{-}_{3'4'})_2(H_2O)_2]^{+}$; Right: $Al[(A^{2-}_{3'4'7})_2(H_2O)_2]^{-}$ complexes.

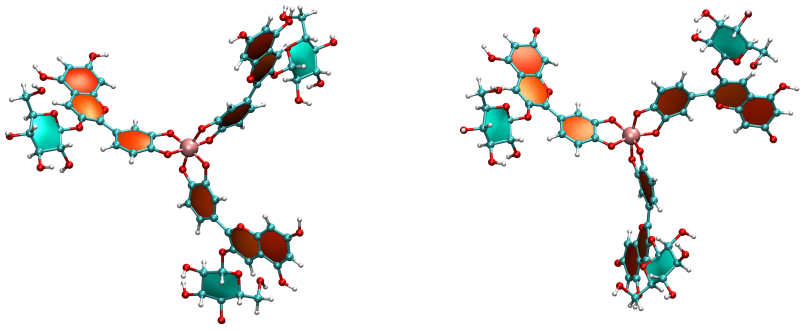


Figure 4.11: Left: $Al[(A^{-}_{3'4'})_3]$; Right: $Al[(A^{2-}_{3'4'7})_3]^{3-}$ complexes.

$Al[(A^{-}_{3'4'})_2(H_2O)_2]^{+}$

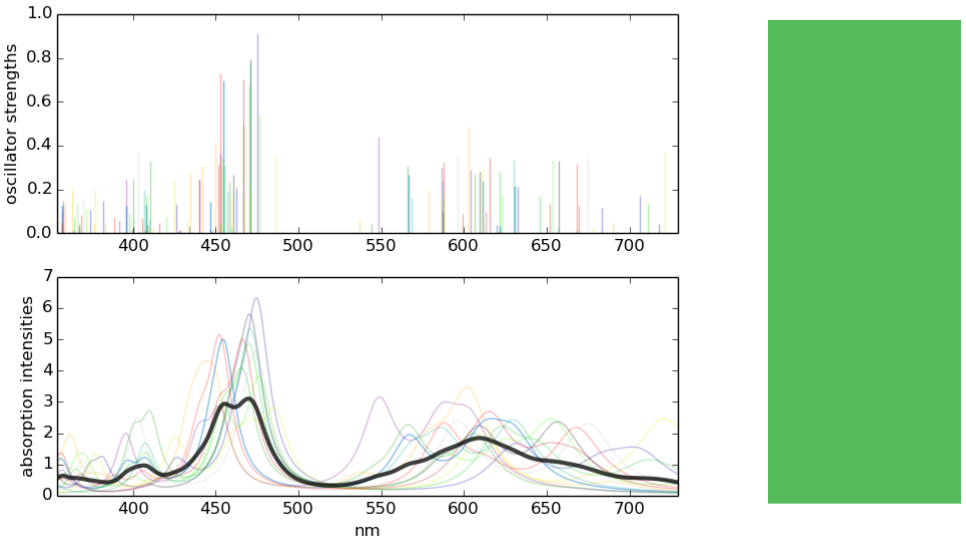


Figure 4.12: Top: transition energies with its corresponding intensities for all the frames used for TDDFT calculations; Bottom: spectrum for each frame and with the bold line the average spectrum over the time; right: the color expressed by the $Al[(A^{-}_{3'4'})_2(H_2O)_2]^{+}$ complex.

$Al[(A^{2-}_{3'4'7})_2(H_2O)_2]^-$

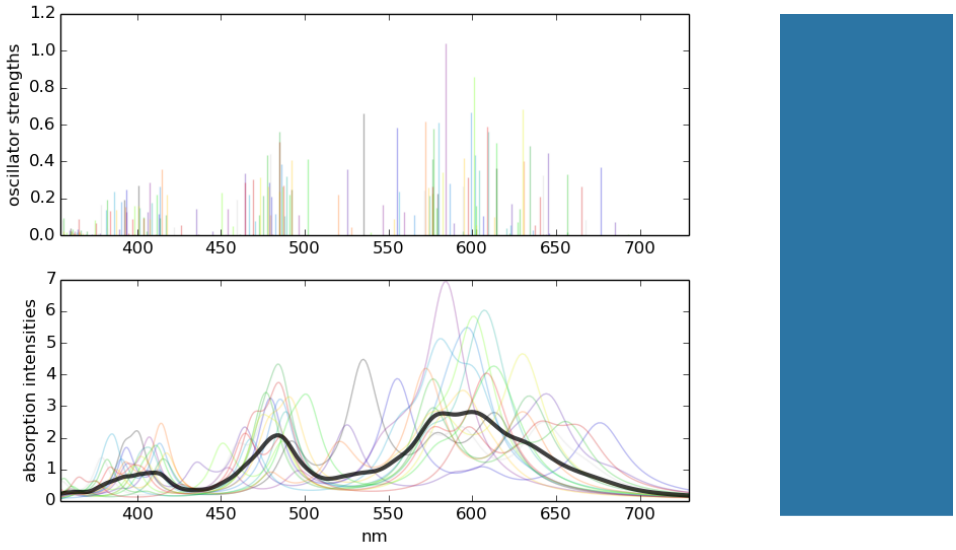


Figure 4.13: Top: transition energies with its corresponding intensities for all the frames used for TDDFT calculations; Bottom: spectrum for each frame and with the bold line the average spectrum over the time; right: the color expressed by the $Al[(A^{2-}_{3'4'7})_2(H_2O)_2]^-$ complex.

$Al[(A^{-}_{3'4'})_3]$

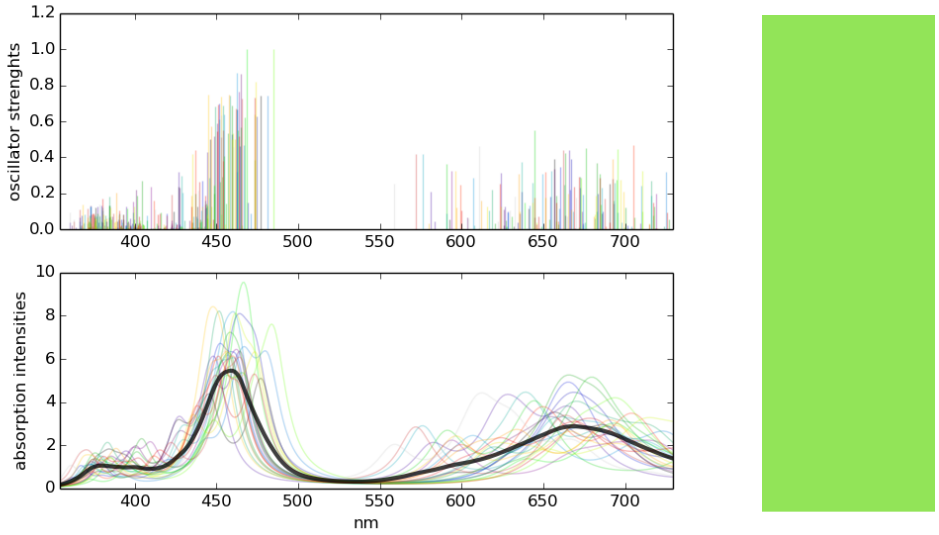


Figure 4.14: Top: transition energies with its corresponding intensities for all the frames used for TDDFT calculations; Bottom: spectrum for each frame and with the bold line the average spectrum over the time; right: the color expressed by the $Al[(A^{-}_{3'4'})_3]$ complex.

$Al[(A^{2-}_{3'4'7})_3]^{3-}$

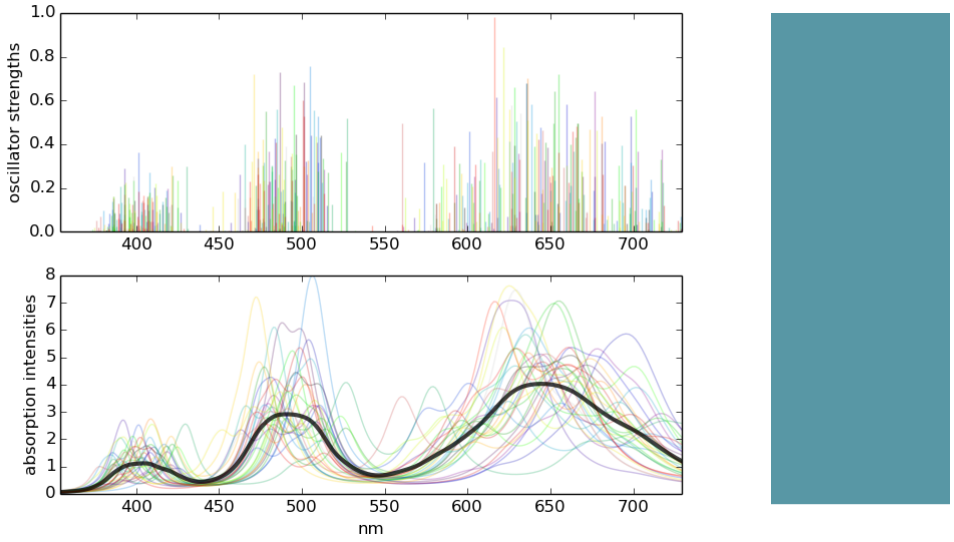


Figure 4.15: Top: transition energies with its corresponding intensities for all the frames used for TDDFT calculations; Bottom: spectrum for each frame and with the bold line the average spectrum over the time; right: the color expressed by the $Al[(A^{2-}_{3'4'7})_3]^{3-}$ complex.

4.4 Beyond experiment: Predictions of the coexisting species at various acidic conditions

So far we have discussed all the anthocyanin metal ion complexes separately, but in solution they coexist with different probabilities, depending on the acidity of the solution and on metal concentration. Due to the agreement of the calculated spectra with the experimental one for pure C3G (chapter 3) we assume that our computational protocol is reliable calculating even for metal complexes spectra in agreement with the experimental ones. By comparing the experimental spectra [87] of C3G and Al^{3+}:C3G (see figure 4.16), and their corresponding colors at different pH conditions to our computed spectra, we attempt to rationalize which are the dominant species existing at distinct pH conditions.

For very acidic aqueous solutions the recorded absorption spectrum of Al^{3+}:C3G is practically indistinguishable from the one of the C3G [81] meaning that there is no effect on absorption spectrum coming from the metal ion. Thus, it is likely that no complexation occurs at pH(1-2) values and the predominant form at those acidic conditions is still the flavilium cation independently from the presence of the metal ions in solution.

At pH 3 the set of peaks corresponding to the flavylium cation disappears, while a new single set of peaks points to occurrence of the first deprotonation, starting at positions 4′ or 7 (see chapter 3) and resulting into $\mathsf{A}^{\circ}_{4'}$ and A°_{7} species. We can easily presume that the small hard Al^{3+} ion can facilitate deprotonation of the chromophore at position 4′ and even at 3′, resulting into $Al[(\mathsf{A}^{-}_{3'4'})(H_2O)_4]^{2+}$ species. So at pH 3-4 combination of $\mathsf{A}^{\circ}_{4'}$, A°_{7} and $Al[(\mathsf{A}^{-}_{3'4'})(H_2O)_4]^{2+}$ would coexist.

Experimental results show that the addition of aluminum to the buffered solutions in the pH ranges (4-5) results in the appearance of new, batochromically shifted bands [73], induced by metal complexation. Hence, this suggests that metal complexes become more abundant than the $\mathsf{A}^{\circ}_{4'}$ and A°_{7} tautomers. Since no significant second peak appears on the absorption spectrum until pH 7 we can conclude that in the range of pH(4-6) 1:1 metal anthocyanin complexes ($Al[(\mathsf{A}^{-}_{3'4'})(H_2O)_4]^{2+}$ and $Al[(\mathsf{A}^{2-}_{3'4'7})(H_2O)_4]^{+}$) dominate. However, a slightly blueish color of solution at pH 6 indicates that 1:2 complexes start forming.

From pH 7 up, a second high energy peak appears in the absorption spectra displaying even an increasing intensity at pH 8 and 9. This is a

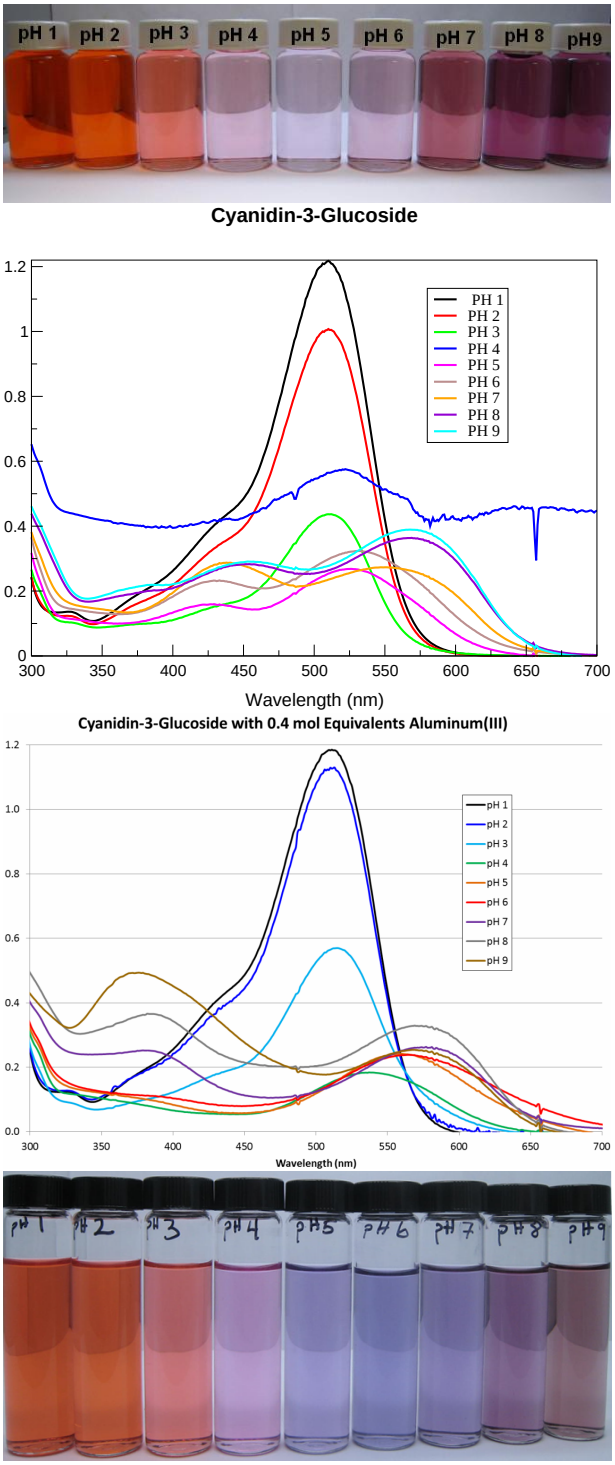


Figure 4.16: First two rows report experimental colors and spectrum at different pH values for cyanidin-3-glucoside in water solution. Third and fourth rows display the experimental spectrum and colors at different pH values for cyanidin-3-glucoside with 0.4 mol equivalent of Al^{3+}.

clear indication that at pH 7-9, 1:2 and 1:3 complexes start to dominate in solution. This prediction agrees with the experimental work presented in reference [75]. Although that study was focusing on delphinidin, and not on C3G, as in the present case, they find that at pH 5.85 the number of delphinidin molecules bonded to one Al^{3+} ion in the complex was around 2.78 at ligand concentration $5X10^{-5}$ mol/L and aluminum concentration $1.8X10^{-5}$ mol/L. Moreover, the maximum concentration of complex was equal to maximum concentration of aluminium. This suggests that at high pH conditions metal ions tend to bind 2 or 3 chromophores. This most probably holds true also for C3G even though the equilibrium of those complexes is shifted from pH 5.85 to pH$\sim$ 7.

4.5 Conclusion

In this study we have investigated the key role that aluminium plays on the absorption spectrum of C3G. Metal complexation appears to affect the ligand's conformational space, thus resulting in changes in the absorption spectrum and expression of bluer or sometimes even greener colors. The pH was found to play an important role on the expression of the pigment chelates, on the extent of the bathochromic shift and, furthermore, on the Al^{3+}:anthocyanin ratio. Our approach was used to qualitatively predict the which species dominate at different pH values by comparing the calculated spectra with the experimental ones.

Chapter 5

Intramolecular Copigmentation

The optical properties of anthocyanins found in nature strongly depend on the presence in solution of other organic compounds and/or metal ions, that [88], are forming molecular complexes and are generating a change in color [89, 90]. While being optically inactive by themselves, when mixed with an anthocyanin solution, interactions with the anthocyanin chromophore generally results in a hyperchromic effect and bathochromic shift in the ebsorption spectra [79]. This phenomenom is known as copigmentation. The co-pigments can be flavonoids, alkaloids, amino acids, organic acids, nucleotides, polysaccharides, metals or another anthocyanins. The most important copigmentation mechanisms are inter- and intramolecular complex formations. The first refers to copigments being co-solvated with the anthocyanin, like it was the case in chapter 4 (1:2 and 1:3 metal:C3G complexes), whereas the latter refers to copigments that are functionally attached to it. In this chapter we will only consider intramolecular copigmentation on a cyanidin backbone.

As pointed out in previous chapters, ordinary anthocyanins are very unstable in slightly acidic or neutral solutions. While copigments by themselves do not contribute to the color, they tend to stabilize the colored species in solution. Thus, when copigments are also present in solution, anthocyanins express stronger colors, therefore the colored forms must be stabilized by some effect of the copigment [1, 5]. In the case where the copigment is a phenolic acid, it will usually folds to surround the anthocyanin, creating hydrophobic interactions between the chromophore and copigment, leading to a structure which is less susceptible a nucleophilic

ANT	Substitution					
	$\mathbf{R}_1$	$\mathbf{R}_2$	Acyl at $\mathbf{R}_1$		Acyl at $\mathbf{R}_2$	
			$\mathbf{R}'_1$	$\mathbf{R}'_2$	$\mathbf{R}'_1$	$\mathbf{R}'_2$
Cyanidin(Cy)-3-sophoroside-5-glucoside	H	H	–	–	–	–
Cy-3-[2-(glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside	Acyl	H	H_3CO	H_3CO	–	–
Cy-3-[2-(2-(sinapoyl)-glucosyl)-glucoside]-5-glucoside	H	Acyl	–	–	H_3CO	H_3CO
Cy-3-[2-(2-(sinapoyl)-glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside	Acyl	Acyl	H_3CO	H_3CO	H_3CO	H_3CO

Table 5.1: Abbreviations of the structures and substitutions patterns of copigmented anthocyanins. For $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}'_1$ and $\mathbf{R}'_2$ see fig. 5.1

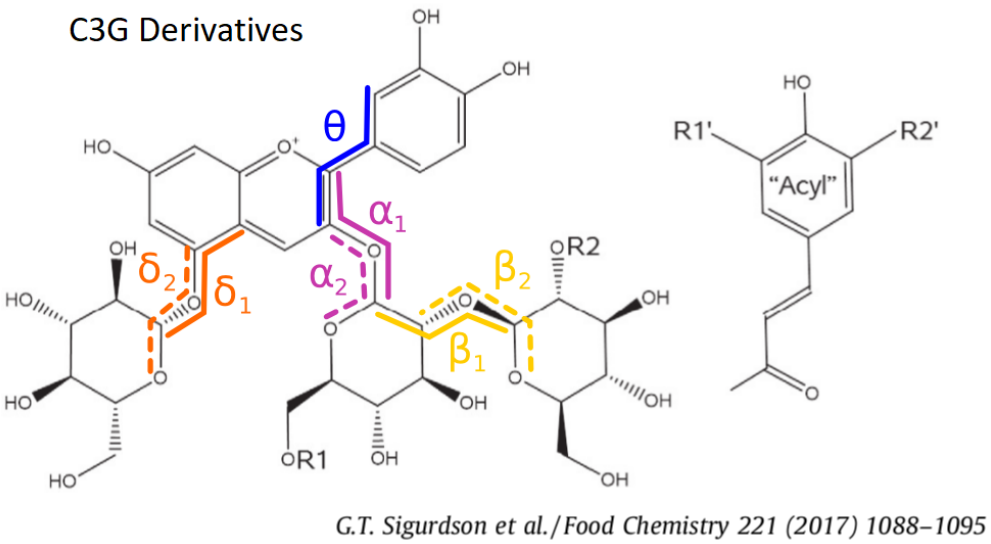


Figure 5.1: Scheme of possible substitution patterns for anthocyanins. Adapted from [5]

attack by water molecules. Thus, co-pigments play a big role in stabilizing anthocyanins. Moreover, they strongly affect the geometry of the pigment they are attached to, hence, affecting optical properties of the whole system.

In our group we have studied 4 different systems [91] (see fig. 5.2 and table 5.1) with different co-pigmentation patterns to study their effect on optical properties. These are all based on the cyanidin backbone and follow a logical progression from the monoglycosylated C3G, extensively studied in the previous chapters:

- The triglycosylated Cyanidin(Cy)-3-sophoroside-5-glucoside (CtG), which like C3G is not acylated, but has additional sugars at positions 3 and 5.
- The triglycosylated, and monoacylated Cy-3-[2-(2-(sinapoyl)-glucosyl)-glucoside]-5-glucoside ($CtGS_1$), which is CtG acylated with a sinapoyl at site R1 (figure 5.1).
- The triglycosylated, and monoacylated Cy-3-[2-(2-(sinapoyl)-glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside ($CtGS_2$), which is CtG acylated with a sinapoyl at site R2 (figure 5.1).
- The triglycosylated, and monoacylated Cy-3-[2-(2-(sinapoyl)-glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside ($CtGS_1S_2$), which is CtG acylated

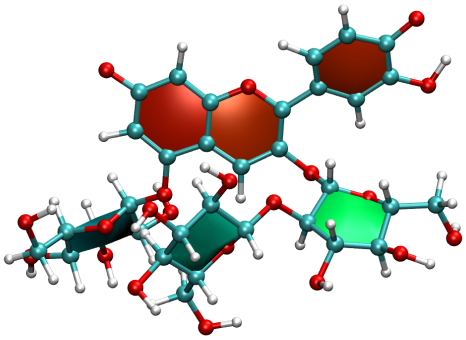
with sinapoyls at both site R1 and R2 (figure 5.1).

The configuration space for each of the molecules was sampled with the protocol discussed in 3.2, with conformational analysis using the clustering algorithm [15] to identify molecular conformers. The choice of parameters for the analysis was made to capture the internal degrees of freedom of the cyanin core, and the possible π-stacking of the aromatic ring of the acyl group(s) with the double ring of the chromophore. The dihedral angles chosen for clustering analysis include the ones highlighted on figure 5.1, as well as the dihedrals along the tail of the phenolic acid, when applicable. Work is still in progress and details are yet unpublished [91].

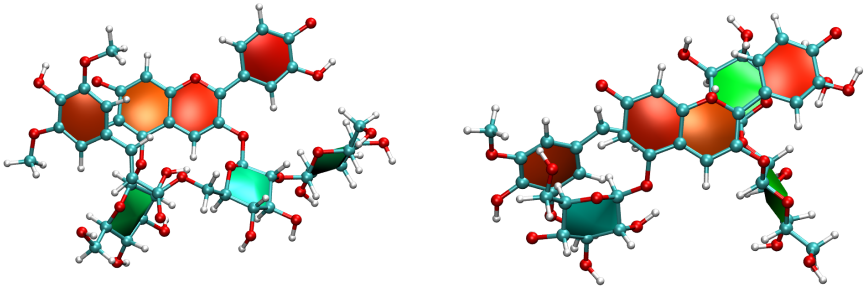
The comformational space of non acylated CtG is very similar to the one of C3G discussed in chapter 3. After clustering analysis, the 2 most populated conformers were selected for further calculations. The protocol discussed in 3.2.1 was applied to both conformers. The corresponding average spectra and their expressed color are shown in figure 5.6. The maximum absorption wavelength, λ_{max}, is determined by the $S_0 \rightarrow S_1$ transition, which is essentially a single-configuration HOMO $\rightarrow$ LUMO transition ($\pi \rightarrow \pi*$, accounting for $\gtrsim 90\%$ of the total oscillator strength) as it was the case of C3G (section 3). The color we predict at neutral pH(4-6) ranges is in good agreement with experimental evidence [92, 93]. In terms of stability, experimental works show [93] that CtG is as unstable as C3G in slightly acidic and neutral pH ranges, thus, we can conclude that glycosylation alone is not enough to protect the chromophore from nucleophilic attack.

When the co-pigment is another phenolic compound like in the cases studied here ($CtGS_1$, $CtGS_2$, and $CtGS_1S_2$), their is a $\pi-$stacking interaction between the acyl groups and the chromophore. According to experiments a bathochromic shift in $\lambda_{\max}$ is observed with respect to the non acylated species, with the largest shift occurring around pH 6 [93]. We expect the negatively charged species to be dominant at these pH and thus will focus our study on this charge state.

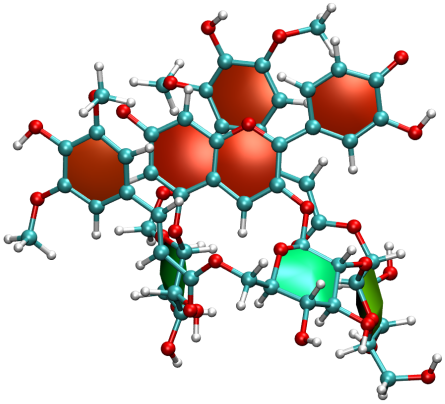
We have studied the conformations of these monoacylated and diacylated systems using enhanced sampling and clustering, as previously for C3G, in each case for the negative species. The most populated conformers for $CtGS_1$ are displayed on the figure 5.3. Clustering Analysis shows that for the most populated conformers π-stacking occurs between the acyl and the A-C rings. We will hereafter refer to these conformer types as "π-stacked". Analysis of the conformational space shows that this type of



Cyanidin-3-sophoroside-5-glucoside (CtG)

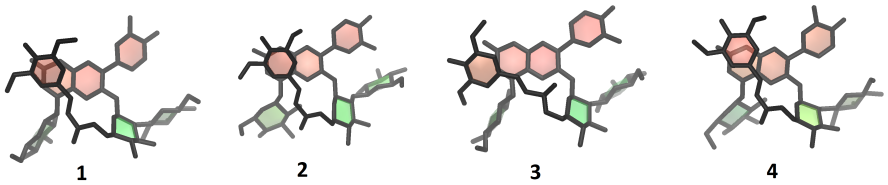


Left: Cy-3-[2-(glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside ($CtGS_1$); Right: Cy-3-[2-(2-(sinapoyl)-glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside ($CtGS_2$)

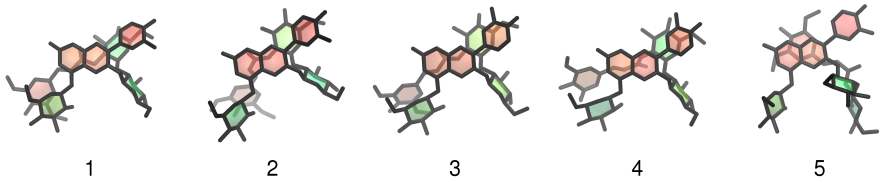


Cy-3-[2-(2-(sinapoyl)-glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside ($CtGS_1S_2$)

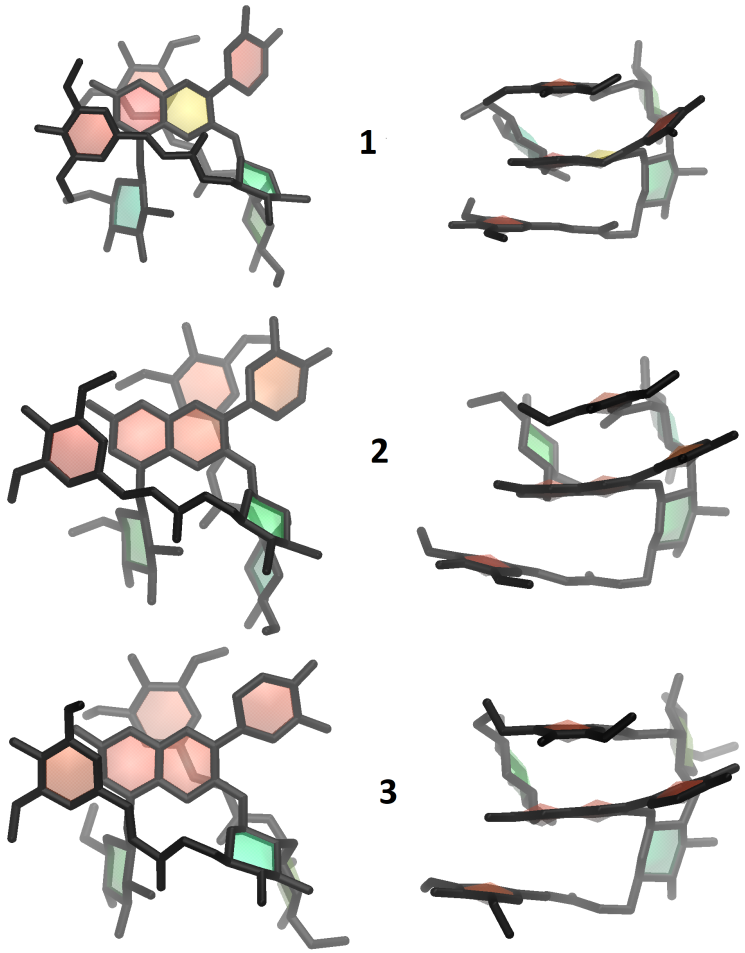
Figure 5.2: The structures of the copigmented anthocyanins studied here.



Cy-3-[2-(glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside.

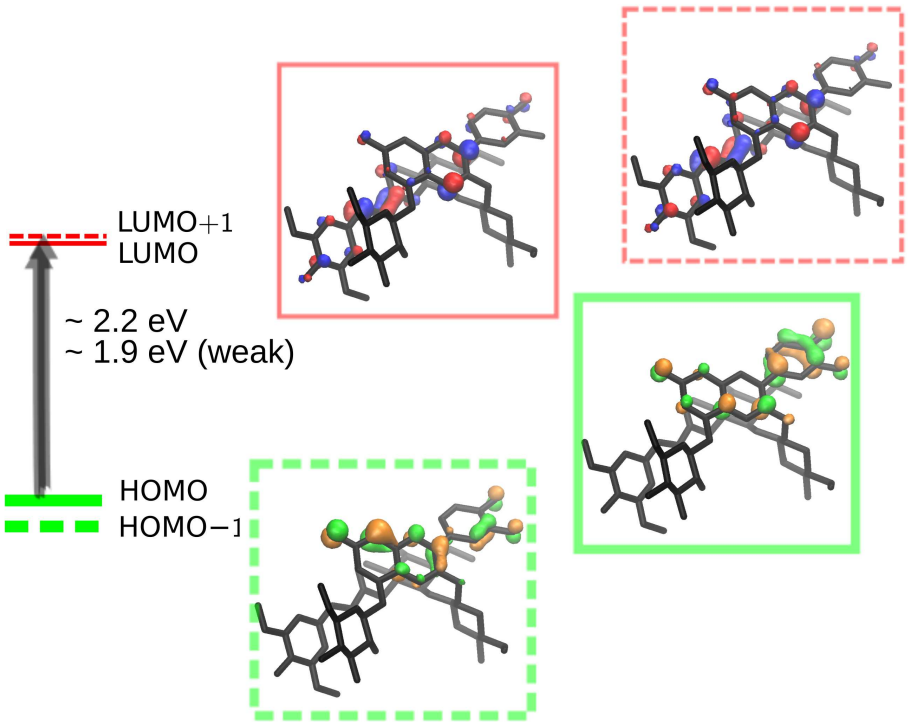


Cy-3-[2-(2-(sinapoyl)-glucosyl)-glucoside]-5-glucoside.



Cy-3-[2-(2-(sinapoyl)-glucosyl)-6-(sinapoyl)-glucoside]-5-glucoside.

Figure 5.3: The most populated conformers coming from clustering analyses, done on 1 μs MD trajectory for each species. For the later molecule top and side view of the clusters are shown for clarity.



Cy-3-[2-(2-(sinapoyl)-glucosyl)-glucoside]-5-glucoside.

Figure 5.4: Diagrammatic overview of frontier orbitals for $CtGS_2$. The HOMO and HOMO-1 orbitals are localized on the cromophore (AC and B rings), whereas the virtual orbitals (LUMO and LUMO+1) are strongly delocalized on chromophore and especially on conjugated chane of the acylic group.

stacking is robust with respect to the orientations of the sugars, and the dihedral θ.

Analysis of $CtGS_2$ shows two main conformer types (different π-stacked arrangments between acyl and anthocyanin). The first 3 conformers show π-stacking between the acyl and the sugar at position 5. Analysis of the conformational space shows that this type of stacking is also robust with respect to the orientations of the sugars, and the dihedral θ. The fourth and fifth conformers exhibit a type of stacking between the acyl ring and the A-C rings. We will refer to these conformer types as "sugar-stacked" and "π-stacked", respectively (see fig 5.3). Analysis of $CtGS_1S_2$ shows conformers that are all π-stacked. We always observe *complete* stacking, where the acyl moieties tend to stay on their preferred side (so called "front" and "back" stacking). Such pronounced stacking reduces conformational

flexibility.

As it has been reported in different studies that acylated anthocyanins exhibit a higher stability than non-acylated ones [93], our analysis further confirms it, since stacking with the acyl group creates a hydrophobic environment around the chromophore, inhibiting the approach of water to C2 for a potential nucleophilic attack. As a result less degradation of the anthocyanin is expected. As for the comparison between mono- and diacylated molecules, our conformational study shows that the diacylated one is more protected by the acyl groups from both sides, meaning that diacylated should be even more stable. This statement is confirmed by the experimental work in ref. [93], reporting a higher pronounced stability of the diacylated pigment at pH 7 and 8 (at these conditions only negatively charged species exist) compare to monoacylated one.

Due the size of the acylated systems, we were able to perform explicit solvent BOMD dynamics only for the most populated conformer of each mono acylated pigment, to have a qualitative picture about the exited states. The same protocol for producing AIMD trajectories and absorption spectrum was used as in the case of metal complexes (discussed in previous chapter). Molecular configurations for TDDFT calculations were extracted from AIMD trajectory at every 0.5 ps, ensuring statistical independence.

Since the acylated molecules have some degree of stacking of the ABC moieties with the aromatic acid(s), more than one low-lying excitation is expected, unlike in the case where we only have HOMO $\rightarrow$ LUMO, as in C3G or CtG.

Mono acylated pigments shows two low-lying transitions. π stacking observed is the responsible for having a noticeable second transition in the red side of the spectrum. The first two transitions are a combination of HOMO $\rightarrow$ LUMO and HOMO $\rightarrow$ LUMO+1. HOMO is localized on the ABC rings. LUMO and LUMO+1 are both shared between ABC rings and acid (see fig. 5.4). Both these virtual orbitals (LUMO, LUMO+1) show a pronounced delocalization. These features, enhanced by the π stacking, translate into the overall red shift of the absorption profile.

The average spectrum for both mono acylated anthocyanins are shown in figure 5.6. According to experimental works [92], a bathochromic shift is expected in acylated anthocyanins compared with non acylated ones. This is confirmed by our results, since C3G and CtG have absorption peaks in the range of 540-560 nm at high pH ranges and acylated anthocyanins have their λ_{max} between 575-585 nm. We still do not have results for the

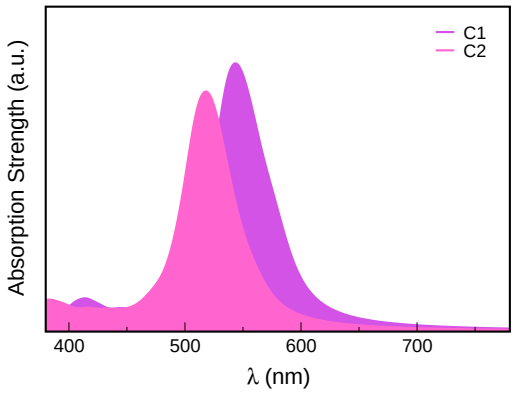
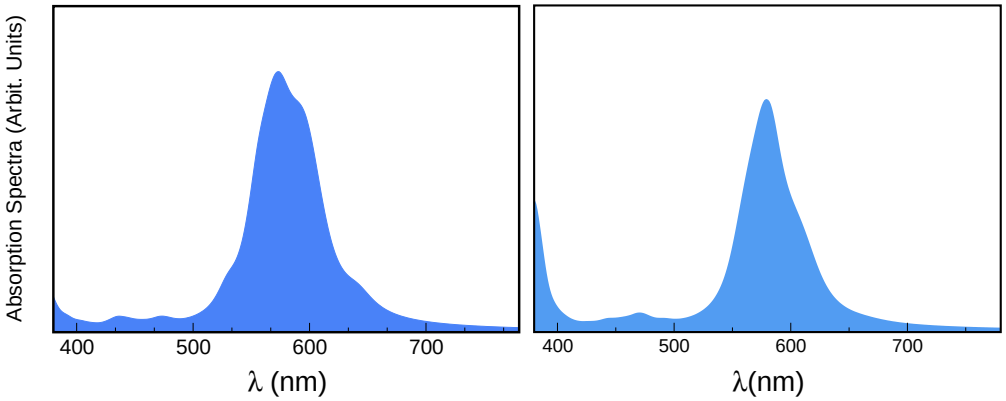


Figure 5.5: CtG, spectrum of two most populated conformers, computed as average of spectra computed on frames extracted from AIMD trajectory. The filling color corresponds to the one each species have in solution.



Left: $CtGS_1$; Right: $CtGS_2$

Figure 5.6: The average spectrum computed on frames extracted from AIMD dynamics. The color refers to that of the pigment in water solution.

diacylated anthocyanin due to its very complicated electronic structure, but we expect an even higher bathochromic shift and blueing effect.

Chapter 6

Conclusion

In this thesis we have introduced a new multiscale computational protocol to simulate the optical spectra of molecular systems in complex environments. We have applied it to study the color optical properties of common natural molecular dyes called anthocyanins, with and without copigments, in water solution. We have highlighted the importance of properly accounting for the configurational flexibility of the molecule, thermal fluctuations, and explicit solvent molecules. Application of this protocol results in excellent agreement between the simulated and experimental spectra, which have been extensively measured by our collaborators for the first time, in support of this work. The quality of this agreement shows that our simulations capture the molecular mechanisms determining the optical properties of C3G in solution, through the description the 5 molecular species shown in figure 3.1, including different charge and protomeric states, thus encompassing the whole pH range.

Future work will be devoted to studying the effects of intermolecular copigmentations and dimerisation, which are important sources of the color diversity of anthocyanins. The main challenge ahead of us is the size of above mentioned systems, which would involve several anthocyanin molecule, as well as copigments, in a single simulation. We need to incorporate our newly developed scheme in a way that we can afford the computation of the absorption spectrum of such systems.

Bibliography

[1] Patrick Trouillas, Juan C. Sancho-GarcÃa, Victor De Freitas, Johannes Gierschner, Michal Otyepka, and Olivier Dangles. Stabilizing and modulating color by copigmentation: Insights from theory and experiment. *Chem. Rev.*, 116(9)(1):4937–4982, 2016.

[2] Raymond Brouillard. Chapter 1 - chemical structure of anthocyanins. In Pericles Markakis, editor, *Anthocyanins As Food Colors*, pages 1 – 40. Academic Press, 1982.

[3] Gregory T. Sigurdson and M. Monica Giusti. Bathochromic and hyperchromic effects of aluminum salt complexation by anthocyanins from edible sources for blue color development. *J. Agric. Food Chem.*, 62(29)(29):6955–6965, 2014.

[4] K. Yoshida, M Mori, and T. Kondo. Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Nat Prod Rep.*, 26(7)(7):884–915, 2009.

[5] G.T. Sigurdson, R.J. Robbins, T.M. Collins, and M.M. Giusti. Spectral and colorimetric characteristics of metal chelates of acylated cyanidin derivatives. *Food Chemistry*, 221:1088 – 1095, 2017.

[6] G. Rastelli, L. Costantino, and A. Albasini. Physico-chemical properties of anthocyanidins. part 1. theoretical evaluation of the stability of the neutral and anionic tautomeric forms. *Journal of Molecular Structure: THEOCHEM*, 279:157 – 166, 1993.

[7] Grace K. Pereira, Paulo M. Donate, and SÃ©rgio E. Galembeck. Effects of substitution for hydroxyl in the b-ring of the flavylium cation. *Journal of Molecular Structure: THEOCHEM*, 392:169 – 179, 1997.

[8] Grace K Pereira and SÃ©rgio E Galembeck. Computational study of the electronic excitations of some anthocyanidins. *Spectrochimica*

Acta Part A: Molecular and Biomolecular Spectroscopy, 54(2):339 – 348, 1998.

[9] Kjell Torskangerpoll, Knut J Borve, Oyvind M Andersen, and Leif J SÃ¦thre. Color and substitution pattern in anthocyanidins. a combined quantum chemical chemometrical study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 55(4):761 – 771, 1999.

[10] Ken Sakata, Norio Saito, and Toshio Honda. Ab initio study of molecular structures and excited states in anthocyanidins. *Tetrahedron*, 62(15):3721 – 3731, 2006.

[11] Araceli Castaneda-Ovando, Ma. de Lourdes Pacheco-Hernandez, Ma. Elena Paez-Hernandez, Josa A. RodrÃguez, and Carlos AndrÃ©s Galan-Vidal. Chemical studies of anthocyanins: A review. *Food Chemistry*, 113(4):859 – 871, 2009.

[12] F.J. Francis. Chapter 6 - stability of anthocyanins in foods. In Pericles Markakis, editor, *Anthocyanins As Food Colors*, pages 181 – 207. Academic Press, 1982.

[13] Mary J. Sever and Jonathan J. Wilker. Visible absorption spectra of metal catecholate and metal tironate complexes. *Dalton Trans.*, pages 1061–1072, 2004.

[14] Giovanni Bussi. Hamiltonian replica exchange in gromacs: a flexible implementation. *Molecular Physics*, 112(3-4):379–384, 2014.

[15] Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

[16] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, Nov 1985.

[17] Iurii Timrov, Marco Micciarelli, Marta Rosa, Arrigo Calzolari, and Stefano Baroni. Multimodel approach to the optical properties of molecular dyes in solution. *Journal of Chemical Theory and Computation*, 12(9):4423–4429, 2016. PMID: 27442613.

[18] D. Frenkel and B. Smit. *Understanding Molecular Simulation*. Academic Press, 2002.

[19] M. P. Allen and D. J. Tildesley. *Computer Simulations of Liquids*, volume 195. Clarendon Press, 1989.

[20] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts.

[21] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[22] Peter. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, 93(7):2395–2417, 1993.

[23] Cameron Abrams and Giovanni Bussi. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*, 16(1):163–199, 2014.

[24] Glenn M. Torrie and John P. Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters*, 28(4):578 – 581, 1974.

[25] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187 – 199, 1977.

[26] Mihaly Mezei. Adaptive umbrella sampling: Self-consistent determination of the non-boltzmann bias. *Journal of Computational Physics*, 68(1):237 – 248, 1987.

[27] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[28] Giovanni Bussi, Alessandro Laio, and Michele Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.*

[29] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, Jan 2008.

[30] Parrinello Michele Barducci Alessandro, Bonomi Massimiliano. Metadynamics. *WIREs Comput Mol Sci*, pages 826 – 843, 2011.

[31] Robert H. Swendsen and Jian-Sheng Wang. Replica monte carlo simulation of spin-glasses. *Phys. Rev. Lett.*, 57:2607–2609, Nov 1986.

[32] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[33] E. Marinari and G. Parisi. Simulated tempering: A new monte carlo scheme. *EPL (Europhysics Letters)*, 19(6):451, 1992.

[34] Ulrich H.E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1):140 – 150, 1997.

[35] Yuji Sugita and Yuko Okamoto. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chemical Physics Letters*, 329(3):261 – 270, 2000.

[36] Mahmoud Moradi, Volodymyr Babin, Celeste Sagui, and Christopher Roland. *Recipes for Free Energy Calculations in Biomolecular Systems*, pages 313–337. Humana Press, Totowa, NJ, 2013.

[37] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1):141 – 151, 1999.

[38] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, 113(15):6042–6051, 2000.

[39] Pu Liu, Byungchan Kim, Richard A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences*, 102(39):13749–13754, 2005.

[40] Lingle Wang, Richard A. Friesner, and B. J. Berne. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (rest2). *The Journal of Physical Chemistry B*, 115(30):9431–9438, 2011. PMID: 21714551.

[41] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008. PMID: 26620784.

[42] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[43] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[44] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52:997–1000, Mar 1984.

[45] Delano P. Chong. *FRONT MATTER*, pages i–xii.

[46] Mark E. Casida. Propagator corrections to adiabatic time-dependent density-functional theory linear response theory. *The Journal of Chemical Physics*, 122(5):054111, 2005.

[47] Dominik Marx and JÃ¼rg Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[48] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, 1994.

[49] F.J. Francis. Chapter 5 - distribution of anthocyanins in food plants. In Pericles Markakis, editor, *Anthocyanins As Food Colors*, pages 181 – 207. Academic Press, 1982.

[50] Ivo Cacelli, Alessandro Ferretti, and Giacomo Prampolini. Predicting light absorption properties of anthocyanidins in solution: a multi-level computational approach. *Theoretical Chemistry Accounts*, 135(6):156, Jun 2016.

[51] Vincenzo Barone, Alessandro Ferretti, and Ilaria Pino. Absorption spectra of natural pigments as sensitizers in solar cells by td-dft and mrpt2: protonated cyanidin. *Phys. Chem. Chem. Phys.*, 14:16130–16137, 2012.

[52] Toshio Goto and Tadao Kondo. Structure and molecular stacking of anthocyanins flower variation. *Angewandte Chemie International Edition in English*, 30(1):17–33.

[53] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian-09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

[54] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[55] Singh U. Chandra and Kollman Peter A. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2):129–145.

[56] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2), 1983.

[57]

[58] Fernando Pina, Maria J. Melo, Cesar A. T. Laia, A. Jorge Parola, and Joao C. Lima. Chemistry and applications of flavylium compounds: a handful of colours. *Chem. Soc. Rev.*, 41:869–908, 2012.

[59] D. Marx and J. Hutter. Ab initio molecular dynamics: basic theory and advanced methods. *Cambridge, UK: Cambridge University Press*, 2009.

[60] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L

Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, 2009.

[61] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio Jr, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, cükbenli E Kü M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V Nguyen, A Otero de-la Roza, L Paulatto, S PoncÃ©, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum espresso. *Journal of Physics: Condensed Matter*, 29(46):465901, 2017.

[62] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[63] Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys.*, 31(3):1695–1697, 1985.

[64] Oliviero Andreussi, Ismaila Dabo, and Nicola Marzari. Revised self-consistent continuum solvation in electronic-structure calculations. *The Journal of Chemical Physics*, 136(6):064102, 2012.

[65] I. Timrov, O. Andreussi, A. Biancardi, N. Marzari, and S. Baroni. Self-consistent continuum solvation for optical absorption of complex molecular systems in solution. *J. Chem. Phys.*, 142(3), 2015.

[66] K. Anton Feenstra, Berk Hess, and Herman J. C. Berendsen. Improving efficiency of large time-scale molecular dynamics simula-

tions of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999.

[67] R. W. G. Hunt and M. R. Pointer. *Measuring Colour*. Wiley, Chichester, 2011.

[68] Sara Laporte and et al. in preparation, 2018.

[69] Changru Ma, Layla Martin-Samos, Stefano Fabris, Alessandro Laio, and Simone Piccinin. Qmmmw: A wrapper for qm/mm simulations with quantum espresso and lammps. *Computer Physics Communications*, 195:191 – 198, 2015.

[70] F.J. Francis. Chapter 7 - analysis of anthocyanins. In Pericles Markakis, editor, *Anthocyanins As Food Colors*, pages 181 – 207. Academic Press, 1982.

[71] S. Asen, R.N. Stewart, and K.H. Norris. Co-pigmentation of anthocyanins in plant tissues and its effect on color. *Phytochemistry*, 11(3):1139 – 1144, 1972.

[72] Laura Estévez, Nicolás Otero, and Ricardo A. Mosquera. Molecular structure of cyanidin metal complexes: Al(iii) versus mg(ii). *Theoretical Chemistry Accounts*, 128(4):485–495, Mar 2011.

[73] Jasmina M. Dimitric Markovic, Dragan S. Veselinovic, Jelisaveta M. Baranac, and Tanja P. Brdaric. Spectroscopic and theoretical study of cyanidin aluminum (iii) complexes. *Spectroscopy Letters*, 41(3):104–115, 2008.

[74] Volodymyr S. Fedenko, Sergiy A. Shemet, and Marco Landi. Uv vis spectroscopy and colorimetric models for detecting anthocyanin-metal complexes in plants: An overview of in vitro and in vivo techniques. *Journal of Plant Physiology*, 212:13 – 28, 2017.

[75] Dragan S. Veselinovic, Jelisaveta M. Baranac, Zoran D. Zujovic, and Dragana S. Djordjevic. Spectroabsorptiometric investigations of complexing reactions of polyhydroxylic flavylium compounds. *Journal of Agricultural and Food Chemistry*, 40(12):2337–2340, 1992.

[76] Henry D. Schreiber, Amy M. Swink, and Taylor D. Godsey. The chemical mechanism for al3+ complexing with delphinidin: A model for

the bluing of hydrangea sepals. *Journal of Inorganic Biochemistry*, 104(7):732 – 739, 2010.

[77] Gregory T. Sigurdson and M. Monica Giusti. Bathochromic and hyperchromic effects of aluminum salt complexation by anthocyanins from edible sources for blue color development. *Journal of Agricultural and Food Chemistry*, 62(29):6955–6965, 2014. PMID: 24547952.

[78] G.T. Sigurdson, R.J. Robbins, T.M. Collins, and M.M. Giusti. Spectral and colorimetric characteristics of metal chelates of acylated cyanidin derivatives. *Food Chemistry*, 221:1088 – 1095, 2017.

[79] G.T. Sigurdson, R.J. Robbins, T.M. Collins, and M.M. Giusti. Evaluating the role of metal ions in the bathochromic and hyperchromic responses of cyanidin derivatives in acidic and alkaline ph. *Food Chemistry*, 208:26 – 34, 2016.

[80] Jun Ren, Sheng Meng, Ch. E. Lekka, and Efthimios Kaxiras. Complexation of flavonoids with iron. structure and optical signatures. *The Journal of Physical Chemistry B*, 112(6):1845–1850, 2008. PMID: 18211058.

[81] Margarida C Moncada, Sara Moura, M.JoÃ£o Melo, Ana Roque, Carlos Lodeiro, and Fernando Pina. Complexation of aluminum(iii) by anthocyanins and synthetic flavylium salts: A source for blue and purple color. *Inorganica Chimica Acta*, 356:51 – 61, 2003. Protagonists in Chemistry, Frausto Da Silva.

[82] By GERALD LIPPERT, JURG HUTTER, and MICHELE PARRINELLO. A hybrid gaussian and plane wave density functional scheme. *Molecular Physics*, 92(3):477–488, 1997.

[83] Giovanni Bussi and Michele Parrinello. Stochastic thermostats: comparison of local and global schemes. *Computer Physics Communications*, 179(1):26 – 29, 2008. Special issue based on the Conference on Computational Physics 2007.

[84] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[85] Giovanni Bussi, Tatyana Zykova-Timan, and Michele Parrinello. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics*, 130(7):074101, 2009.

[86] Paolo Raiteri, Julian D Gale, and Giovanni Bussi. Reactive force field simulation of proton diffusion in bazro 3 using an empirical valence bond approach. *Journal of Physics: Condensed Matter*, 23(33):334213, 2011.

[87] Rebecca J. Robbins, Thomas M. Collins, and M. Monica Giusti. Private collaboration.

[88] Araceli Castaneda Ovando, Ma. de Lourdes Pacheco-Hernandez, Ma. Elena Paez Hernandez, Jose A. Rodriguez, and Carlos Andres Galan Vidal. Chemical studies of anthocyanins: A review. *Food Chemistry*, 113(4):859 – 871, 2009.

[89] Mirela Kopjar et al. Copigmentation effect of phenolic compounds on red currant juice anthocyanins during storage. 2010.

[90] Rodrigo N. Cavalcanti, Diego T. Santos, and Maria Angela A. Meireles. Non-thermal stabilization mechanisms of anthocyanins in model and food systems an overview. *Food Research International*, 44(2):499 – 509, 2011.

[91] S. Laporte et. al. In preparation.

[92] G.T. Sigurdson, R.J. Robbins, T.M. Collins, and M.M. Giusti. Spectral and colorimetric characteristics of metal chelates of acylated cyanidin derivatives. *Food Chemistry*, 221:1088 – 1095, 2017.

[93] Neda Ahmadiani, Rebecca J. Robbins, Thomas M. Collins, and M. Monica Giusti. Molar absorptivity and spectral characteristics of cyanidin-based anthocyanins from red cabbage. *Food Chemistry*, 197:900 – 906, 2016.